THE UNIVERSITY *of* EDINBURGH

Protein-directed Dynamic Combinatorial Chemistry

A thesis presented for the degree of

DOCTOR OF PHILOSOPHY IN

ORGANIC CHEMISTRY

by

Venugopal T. Bhat

School of Chemistry

The University of Edinburgh

2011

Supervisor: Dr. Michael F. Greaney

Dedicated to my parents, my wife Jyothi

little Ujwala and Chethana….

Declaration

I herewith declare that all of the scientific work and experiments reported in this thesis are my own unless otherwise noted. None of this work has been submitted in any other application for a higher degree.

Venugopal T. Bhat

Table of Contents

Chapter 2 Catalysing acylhydrazone dynamic libraries

List of abbreviations

ACE	acetylcholinesterase
BSA	bovine serum albumin
CA	carbonic anhydrase
CaM	calmodulin
CDNB	1-chloro-2,4-dinitrobenzene
DCC	dynamic combinatorial chemistry
DCL	dynamic combinatorial library
DCM	dichloromethane
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
EA	ethacrynic acid
EDTA	ethylenediaminetetraacetic acid
EPNP	1,2-epoxy-3-(p-nitophenoxy)propane
EtOH	ethanol
GSH	glutathione
GSSG	glutathione disulfide
GST	glutathione S-transferase
HPLC	high pressure liquid chromatography
Hz, MHz	Hertz, megahertz
IC_{50}	concentration of inhibitor required for 50% inhibition
ITC	isothermal titration calorimetry
LCMS	liquid chromatography mass spectrometry
m.p	melting point

MeCN	acetonitrile
MeOH	methanol
MWCO	molecular weight cut-off
NMR	nuclear magnetic resonance
PBS	phosphate buffered saline
PDB	protein data bank
RNA	ribonucleic acid
RT	room temperature
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
TLC	thin layer chromatography
UV	ultraviolet (light)

Preface

Parts of this thesis have been communicated in the literature and have been co-written by the author of this thesis:

A) V. T. Bhat, A. M. Caniard, T. Luksch, R. Brenk, D. J. Campopiano & M. F. Greaney, **Nucleophilic catalysis of acylhydrazone equilibration for protein-directed dynamic covalent chemistry ,** *Nature Chemistry* **2010**, 2, 490-497

B) M. F. Greaney and V. T. Bhat, **Protein-directed Dynamic Combinatorial Chemistry,** in *Dynamic Combinatorial Chemistry***;** Ed. Miller, B. L.; Wiley, **2009.** (invited contribution)

Acknowledgements

First of all, I would like to thank my supervisor Dr. Michael F. Greaney for giving me the opportunity to work with him on this exciting project. It has been a great learning experience for me which I will cherish for a long time to come. Thank you Mike, for your advice and support, especially during the difficult times.

Many thanks to our collaborators, Dr. Dominic Campopiano and Anne Caniard, without whom this project would not have gone very far.

I would like to thank the Greaney group for the stimulating and friendly atmosphere in the labs. My deepest gratitude to Dr.Stephan Ohnmacht and Dr. Aileen Mitchell for being such great friends. Thank you for your support.

Special thanks to Alex Clipson, my chemical biology mate in the group, for the many helpful discussions we had on our projects.

I would also like to thank all the members of staff at the School of Chemistry, your help and support was extremely valuable.

Finally, to the special people in my life; my parents A.Thrivikrama Bhat and Vasantha, thank you so much for all the love and support. Thank you, Jyothi, for your love and understanding during these tough years. This thesis is dedicated to you and our little Ujwala.

Abstract

Dynamic combinatorial chemistry (DCC) is a novel approach to medicinal chemistry which integrates the synthesis and screening of small molecule libraries into a single step. The concept uses reversible chemical reactions to present a dynamic library of candidate structures to a template which selects and removes the best binder from equilibrium. Using this evolutionary process with a biopolymer template, such as a protein, leads to the protein directing the synthesis of its own best ligand. Biological DCC applications are extremely challenging since the thermodynamic criterion of reversibility has to be met under physiological conditions to ensure stability of the biomolecular template. The list of reversible reactions satisfying these stringent criteria is limited and is a major constraint on achieving both reaction and structural diversity in adaptive dynamic libraries. This thesis reports the development of a catalysed version of acylhydrazone dynamic libraries which are truly adaptive under protein-friendly conditions.

In the presence of aniline as a trans-imination catalyst, acylhydrazone dynamic combinatorial libraries equilibrate rapidly at pH 6.2 and are switched off by an increase in pH. We designed acylhydrazone libraries targeting the enzyme superfamily Glutathione-S-Transferase (GST) using a scaffold aldehyde, 4-chloro-3-nitrobenzaldehyde, which is structurally related to a known GST substrate chlorodinitrobenzene. On interfacing these dynamic libraries with two different GST enzymes (SjGST from the helminth worm *Schistosoma japonicum* and hGSTP1-1, a human isoform and an important oncology drug target) we observed isoform-selective amplification effects with two different acylhydrazones selected by the proteins. To explore the potential of anchoring in our DCC methodology we conjugated the endogenous GST ligand, glutathione (GSH) onto the scaffold chloronitrobenzaldehyde. The GSH recognition motif acts as an anchor and allows us to explore the hydrophobic binding site of the enzyme in a fragment- based approach. The presence of the glutathione moiety led to increased solubility of the library members and a DCC experiment with the enzymes led to the selection of conjugate hydrazones with significant binding ability.

Multi-level dynamic libraries use multiple exchange processes in the same system to increase their accessible structural diversity. These exchange reactions may be *orthogonal*, where the different chemistries can be activated or deactivated independently of each other, or *simultaneous*, where all the processes are dynamic and crossover under the same conditions. Together, these interacting molecular networks provide an exciting experimental approach to the emerging field of systems chemistry. We demonstrate that two reversible reactions, conjugate addition of thiols to enones and hydrazone formation, are fully compatible and orthogonal to one another in a single dynamic library. Hydrazone exchange takes place at acidic pH, while conjugate addition operates at basic pH. Simple pH change can be used to switch between each process and establish two channels of reactivity.

Chapter 1

Dynamic Combinatorial Chemistry

1.1 Introduction

Over the past century synthetic organic chemistry has evolved into a highly mature field with tremendous success in building complex molecular structures, usually in a stepwise fashion. This progress has been aided by an ever-expanding tool-kit of reactions, a vast majority of which operate under kinetic control. These irreversible reactions lead to the formation of stable covalent bonds with fast reaction times and high selectivities. The products of these reactions are determined by the free energy differences among all possible transition states and thus can be controlled by appropriate choice of reaction conditions including catalysts. However, over the past two decades, there has been a renewed interest in chemical reactions carried out reversibly under equilibrium control which led to the emergence of a new discipline: dynamic combinatorial chemistry (DCC).[1,2] This nascent field borrows heavily from supramolecular chemistry and self-assembly and has benefited from the sustained efforts of many well-known research groups throughout the world but the basic conceptual framework was initially established by exemplary work from Lehn and Sanders. [3,4]

Synthetic processes involving DCC combine the robustness of covalent synthesis with the unique features of 'error-checking' and 'proof reading' derived from reversibility.[1] In these systems product formation is under thermodynamic control and the final product distribution at equilibrium depends on the relative ground-state stabilities of the product structures. A very significant advantage of dynamic combinatorial chemistry over the kinetic regime lies in its adaptability: the ability to re-adjust the product distribution as a response to changes in the reaction environment (for example, concentration changes, template effects, pH or temperature) even after the initial products have been formed. These features of DCC have led to applications in a diverse set of areas ranging from materials science (dynamic polymers or dynamers)[5] to sensors[6] and systems chemistry.[7] But the most interesting application of this concept is in drug discovery using nucleic acids or proteins as templates to bias a dynamic combinatorial library (DCL) thus identifying best binding compounds for these drug targets.[8] Though some preliminary publications have appeared in this area, challenges remain, especially in the use of

proteins as templates.[2] Our efforts at meeting some of these challenges form the central theme of this thesis. In the following sections of this introductory chapter we review the design of a DCC experiment and previous literature on the application of DCC and related kinetic approaches to protein targets.

1.2 The design of a DCC experiment

The basic structure of a DCC experiment consists of a library of building blocks which undergo a reversible reaction to generate a molecular population of great diversity at equilibrium.[1,2] A schematic representation of the DCC concept is given in Figure 1.1. Under the particular conditions of the experiment, the composition of the library at equilibrium is determined by the thermodynamic stability of each of the library members. A selection mechanism to bias the library is then implemented which will usually be an added template but may also include other mechanisms (for example, self-selection, physical properties or a phase selection like binding to a target on a solid phase or crystallization).[2]

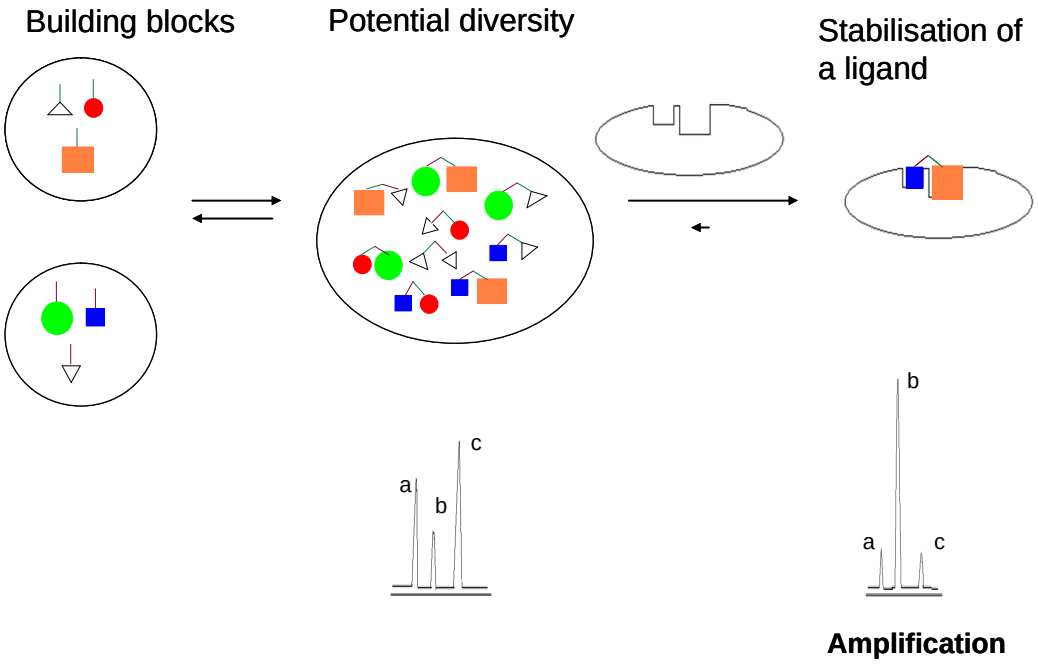
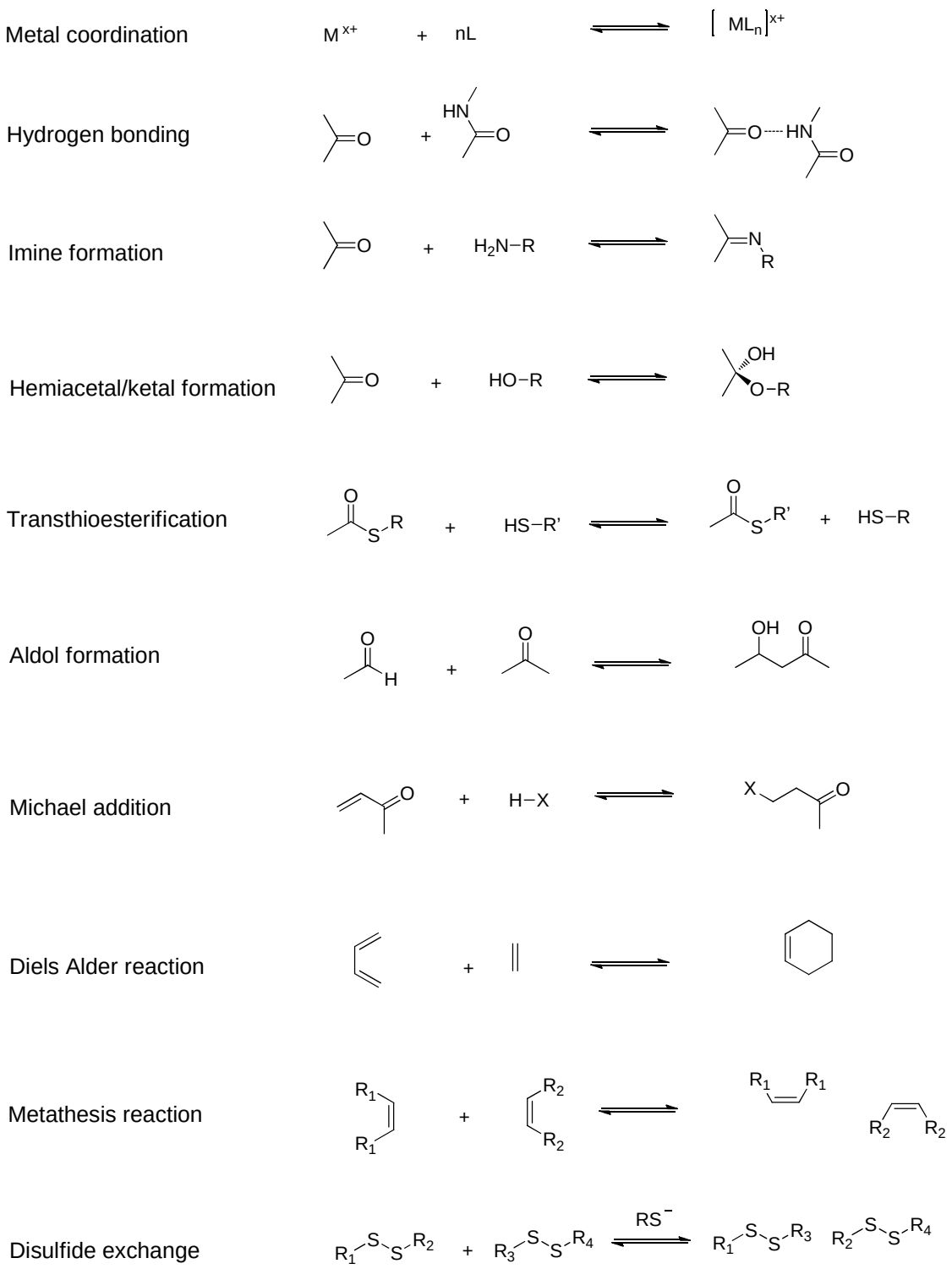


Figure 1.1 *The basic structure of a DCC experiment*

When a template binds tightly to a specific library member, this species is stabilized, usually through non-covalent interactions, and the equilibrium would shift resulting in a higher concentration of the selected member. This relative increase in the concentration of the selected library member is referred to as 'amplification' which can easily be identified by HPLC or LC/MS. Analytical techniques based on separation usually require a 'lock-in' step to fix library composition but can be avoided in the case of *in situ* techniques like NMR or X-ray crystallography.[2]

The key factor responsible for a successful DCC experiment is the choice of the reversible reaction that mediates the exchange of building blocks between the library members.[1-4] An ideal reaction should be reversible on a reasonable timescale under mild conditions and should not interfere with the molecular recognition involved in the selection process. The equilibration rate should be slower than binding with the target. It should also be possible to kinetically freeze the reaction by suitably changing experimental conditions like pH or temperature which helps in the analysis of the DCL composition. The library members synthesized through such a reaction should be soluble at equilibrium; insoluble members can act as thermodynamic or kinetic traps which affect the dynamic nature of the system.[1]

Both covalent and non-covalent bonds (including metal coordination and hydrogen bonds) are useful in dynamic combinatorial chemistry.[1,2] The non-covalent bonds are usually weak and labile leading to fast equilibration times but low thermodynamic stability of the products present analytical difficulties. On the other hand, reversible covalent reactions are slower but form thermodynamically stable products which are usually robust enough for analysis. Examples of reversible reactions used in DCC are listed. (Scheme 1.1)



Scheme 1.1 *Some reversible reactions useful for DCC*

For generating an efficient dynamic combinatorial system the building blocks need to satisfy a few important conditions.[1,2] They should possess appropriate functional groups that take part in the selected reversible reaction. These groups should not interfere with the molecular recognition event during the selection process. Additional solubilising groups or chromophores may be incorporated into the scaffold structure. Ideally all the building blocks should have similar reactivity in order to generate an isoenergetic library which prevents bias towards some components. This condition is practically difficult to meet, given the requirement for generating maximum structural diversity with minimum number of components. A near-isoenergetic library can be generated in some cases by suitably selecting the building block stoichiometry. It is normally expected that the addition of a template to an equilibrating library would selectively amplify the best binding compound but this may not necessarily be the case. Since DCLs tend to lower the overall Gibbs free energy of the system, more than one species may be amplified in some cases. In some other cases the final equilibrium may even favour library members other than the thermodynamically most stable one. Generally, it is possible to optimize experimental conditions to drive the equilibrium exclusively towards the best binding species.

1.3 Protein-directed dynamic combinatorial chemistry

The use of proteins as templates for small molecule DCLs provides a new method for studying protein-ligand interactions.[2] In the DCC process the structure of the protein directs the assembly of its best binder *in situ* thereby combining the synthesis and biological screening of the candidate structures into a single step. In principle this concept is a simple one with the template protein stabilizing one member of the DCL through selective binding and Le Chatelier's principle ensuring its amplification at the expense of other non-binding species. But there are several challenges associated with protein-templated DCC. Most proteins are very sensitive to pH, temperature and a host of chemical reagents. This requires the DCLs to be assembled under strictly physiological conditions which place a severe restraint on the reversible synthetic

chemistry that can be used. Many innovative approaches have been tried to address these issues which are discussed in the following sections with specific examples.

1.3.1 C=N Bond Formation

The first example of using DCC with proteins was reported by Huc and Lehn in 1997.[3,9] They generated an imine DCL with a set of 3 aldehydes **1-3** and 4 amines **a-d** targeting the enzyme bovine carbonic anhydrase II (CA II), a well-characterised Zn (II) metalloenzyme (Figure 1.2). The library components were selected to present structural features similar to those of known CA (II) inhibitors with the 3 aldehydes containing a *para*-substituted sulfonamide motif whose nitrogen atom coordinates to the Zn (II) ion in the enzyme's active site. Substituents at *para*- position to the sulfonamide group interact with a hydrophobic region of the protein above the Zn (II) pocket. In order to generate isoenergetic libraries components with comparable imine-forming reactivity were selected while ensuring activity in the UV-visible spectrum to aid detection and analysis by HPLC.

The DCL was synthesized using a 15-fold excess of amines with respect to each aldehyde scaffold in phosphate buffer (pH 6) in the presence of 3-fold excess of $NaBH_3CN$. The excess amounts of starting primary amines were used to limit side reactions between the aldehydes and the product secondary amines or amino groups of the protein. CA has 18 Lys ε-NH_3^+ groups in addition to the terminal amine. The pseudo-first order conditions established for each aldehyde with every primary amine ensured that there would be no shortage of any amine which could introduce a bias in the library composition. The reducing agent irreversibly reduces the highly unstable imine products of the DCL thus facilitating the isolation and analysis of the mixture at equilibrium.

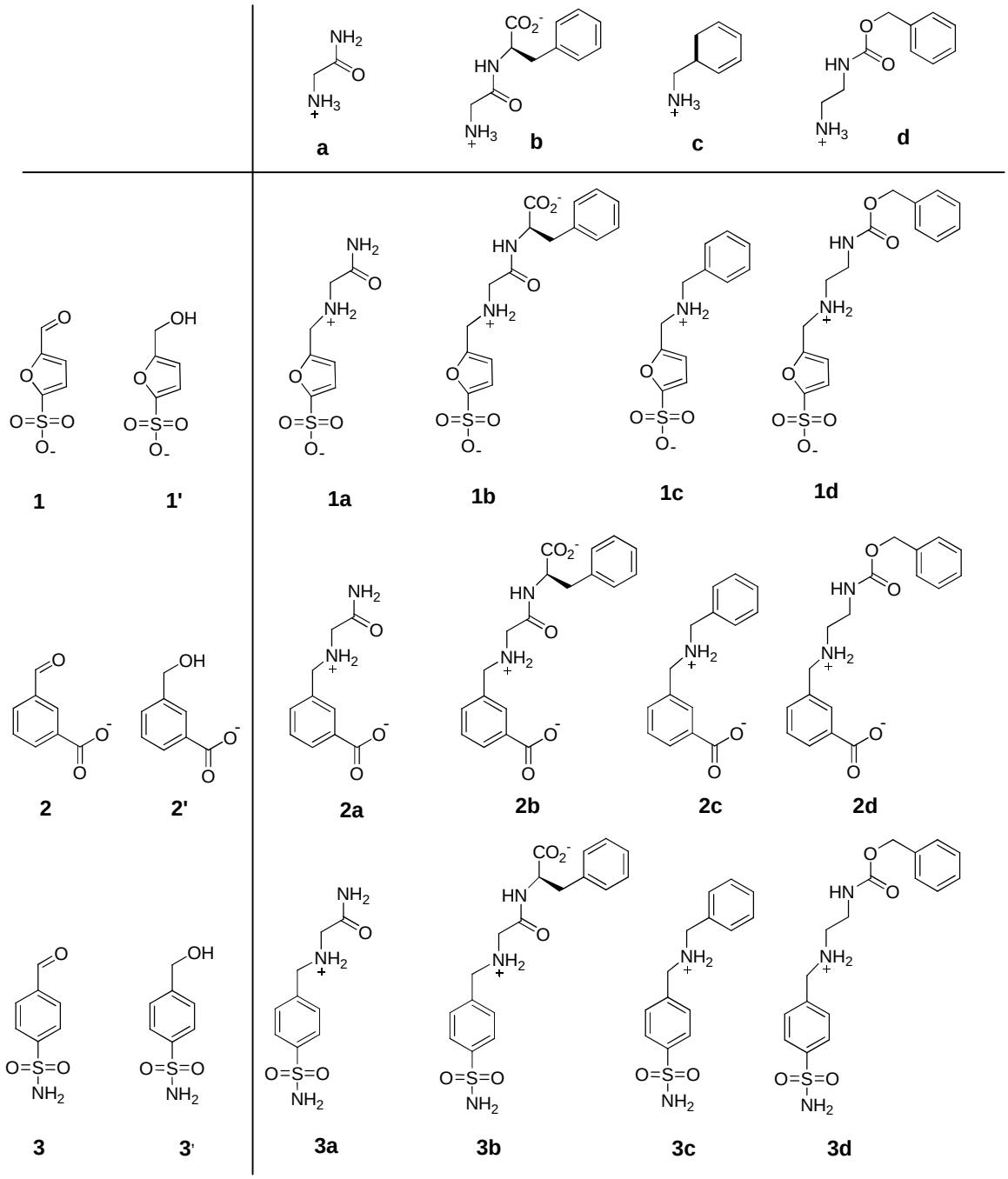


Figure 1.2 *Lehn and Huc's imine DCL for carbonic anhydrase inhibition*

However, a significant drawback of this procedure is that the imine molecules taking part in the molecular recognition events involving the protein are not represented in the final analysis of the DCL. Yet another complicating factor is the possible

decrease in the aldehyde concentration available for imine formation due to its reduction to benzylic alcohols **1'-3'** as side products.

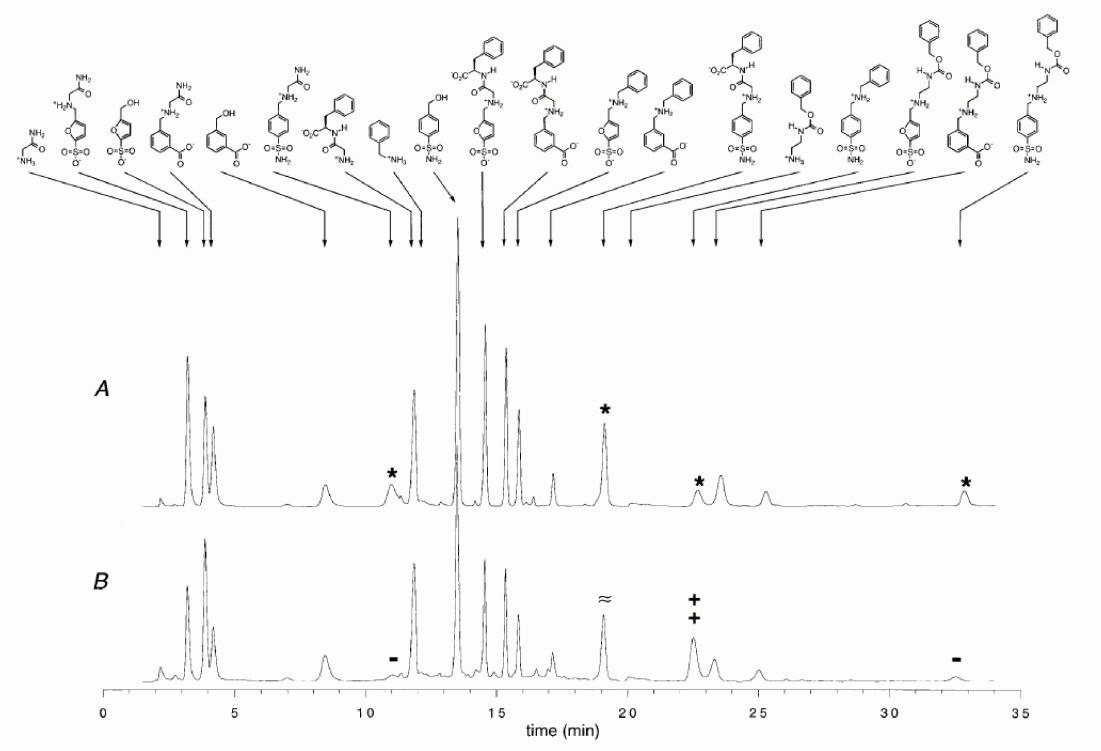


Figure 1.3 *HPLC traces of the reduced imine DCL. Trace A shows the blank DCL composition. Trace B shows the DCl generated in the presence of stoichiometric amounts of the enzyme CA. Amine* ***3c****(++) has been amplified at the expense of* ***3a*** *and* ***3d****(-). Amine* ***3b*** *is relatively unchanged. Reproduced from reference* 3.

A blank DCL as a control was first generated in the absence of the CA template. All the expected imine reduction products were observed in the HPLC trace (Figure 1.3) with equilibrium attained in 24 hours. However, 2 weeks were required for equilibration in the presence of the target protein. A thermal denaturation step was performed to ensure the release of ligands from the enzyme active site. Separation of the denatured enzyme from the reaction products by microcentrifuge filtration preceded HPLC analysis. The templated library composition revealed a number of changes relative to the blank with respect to each set of amines. Overall yields for all reduction products in the presence of the enzyme were lower than the blank. The relative composition of amines derived from aldehydes **1** and **2** were essentially

unchanged but the amount of amines based on the aldehyde scaffold **3** had undergone a change in the presence of the protein. An amplification of the benzylamine derivative **3c** was observed at the expense of **3a** and **3d** while amine **3b** maintained the same relative concentration. A binding series was established by carrying out similar DCL experiments with 4 amines and aldehyde **3** alone. A control DCL synthesized in the presence of a known inhibitor of CA, hexyl 4-sulfamoyl-benzoate, yielded no amplifications indicating that the active site of the protein was responsible for the template effect. The imine precursor of the best binder **3c** selected by the protein is very similar in structure to a known high affinity ligand for CA, 4-sulfamoylbenzoic acid benzylamide (K_d = 1.1 nM). This seminal work by Lehn and Huc established the basic protocol for the construction of imine DCLs targeting biomolecules.

Acyl hydrazone formation for DCL generation in the presence of biological targets was examined by the same research group.[10] The Sanders group had previously used this reaction under thermodynamic control in abiotic systems.[11] The reaction involves aldehyde and hydrazide components forming stable acylhydrazones amenable to analysis. However, a limitation of the reaction for its use with biomolecules is the acidic pH (*ca.* 4) required for fast and effective equilibration. These harsh conditions would denature most proteins and therefore Lehn *et al* adopted a 'pre-equilibrated' DCC approach wherein the DCL is first generated in the absence of any target. Upon equilibration, the library is frozen by raising the pH to neutral conditions and then exposed to the protein target, in this case the enzyme acetylcholine esterase (ACE), in a conventional enzyme assay. Since the protein is introduced only in a second discrete step separate from the DCL synthesis, a wide range of reversible reactions can be used for generating a library with great diversity. However, the adaptive processes that form the cornerstone of DCC are absent in this protocol resulting in a static library of compounds. The second step of the process is similar to the assaying of a conventional combinatorial library and as in these cases a deconvolution strategy is required to identify hit molecules from the mixture.

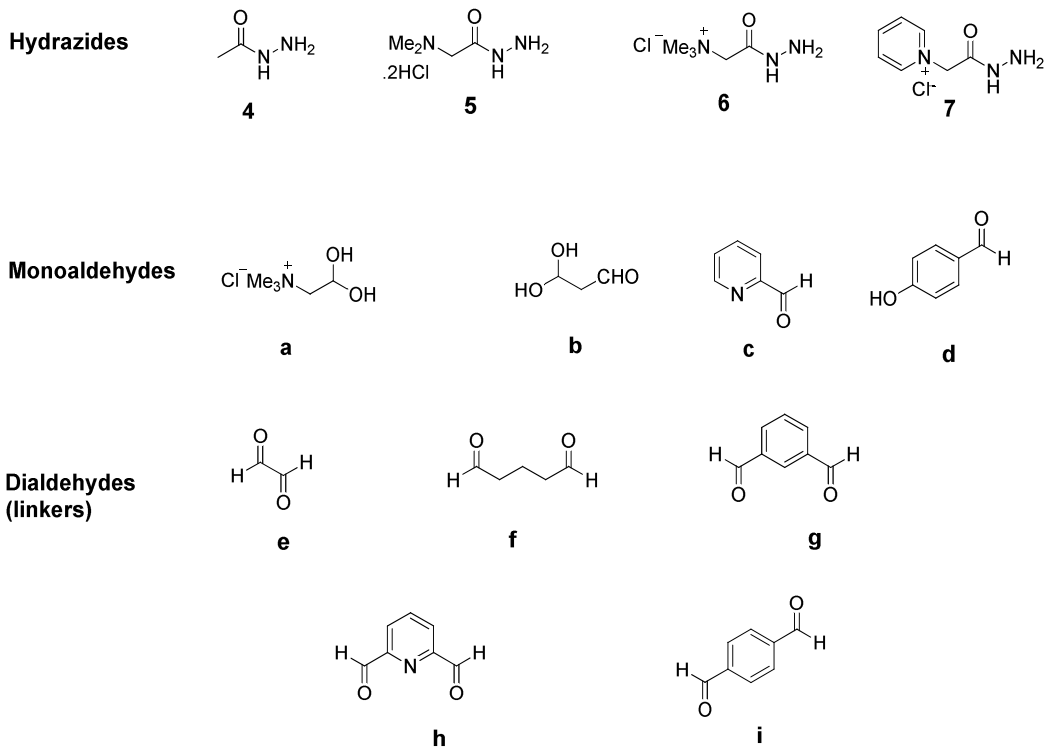


Figure 1.4 *Hydrazides and aldehydes used in the acyl hydrazone DCL*

Lehn and co-workers synthesized all possible acylhydrazones based on the 13 building blocks (Figure 1.4) and assayed their inhibitory activity towards acetylcholine hydrolysis by ACE. To establish a dynamic deconvolution approach they then synthesized a preequilibrated DCL containing all members which was then frozen and assayed. Sub-libraries, each containing all components except a hydrazide or aldehyde building block, were prepared and assayed. Libraries missing either hydrazide **7** or dialdehyde **i**, exhibited an increase in ACE activity thus identifying the bis acyl hydrazone **7-i-7** as the best inhibitor of the enzyme with inhibitory activity in the low nanomolar range.

A similar pre-equilibrated acyl hydrazone DCL method was then used by Lehn *et al* to identify cationic inhibitors of *Bacillus subtilis* HPr kinase.[12] The DCL design was based on the scaffold structure **8**, a previously discovered lead inhibitor (Figure 1.5). A 440-membered DCL using 16 hydrazides containing nitrogen heterocycles, 2 monoaldehydes and 3 dialdehydes was generated under acid equilibration. Again,

based on a dynamic deconvolution strategy using 21 sub-libraries a bis-acyl hydrazone **11** was clearly selected from the available compounds. A subsequent inhibition assay revealed an IC_{50} value of 18 μM.

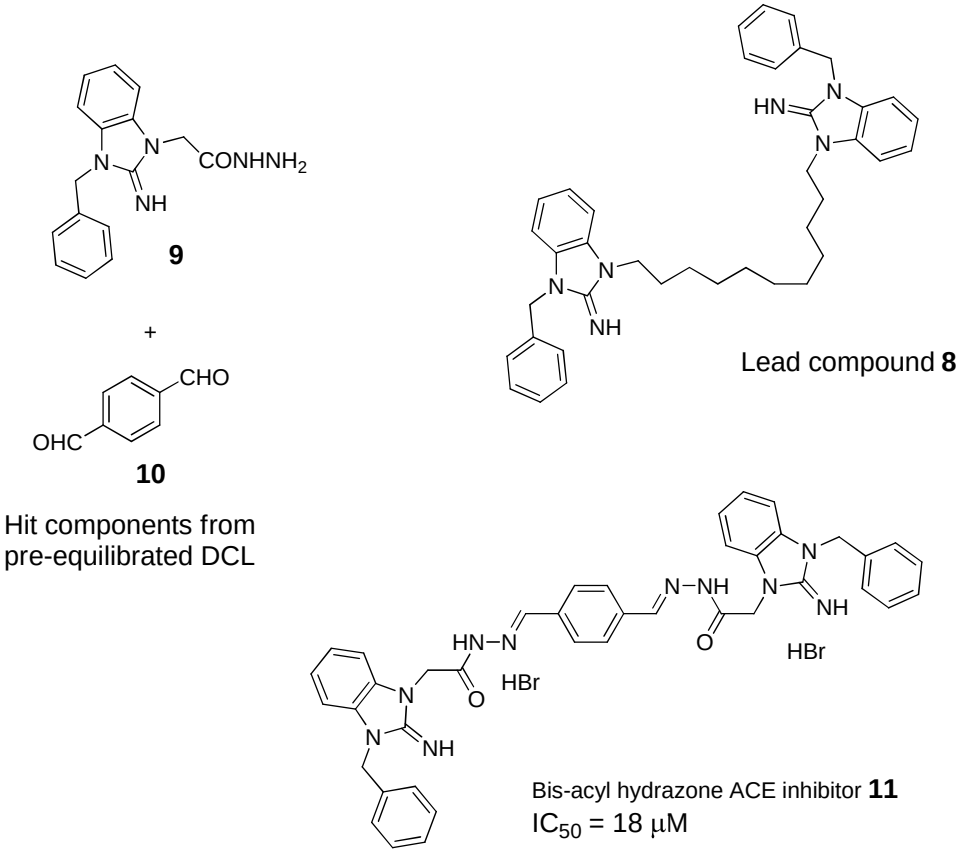
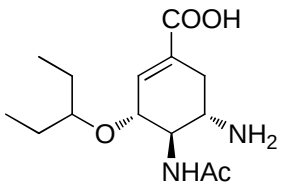


Figure 1.5 *Acyl hydrazone DCL components for kinase inhibition*

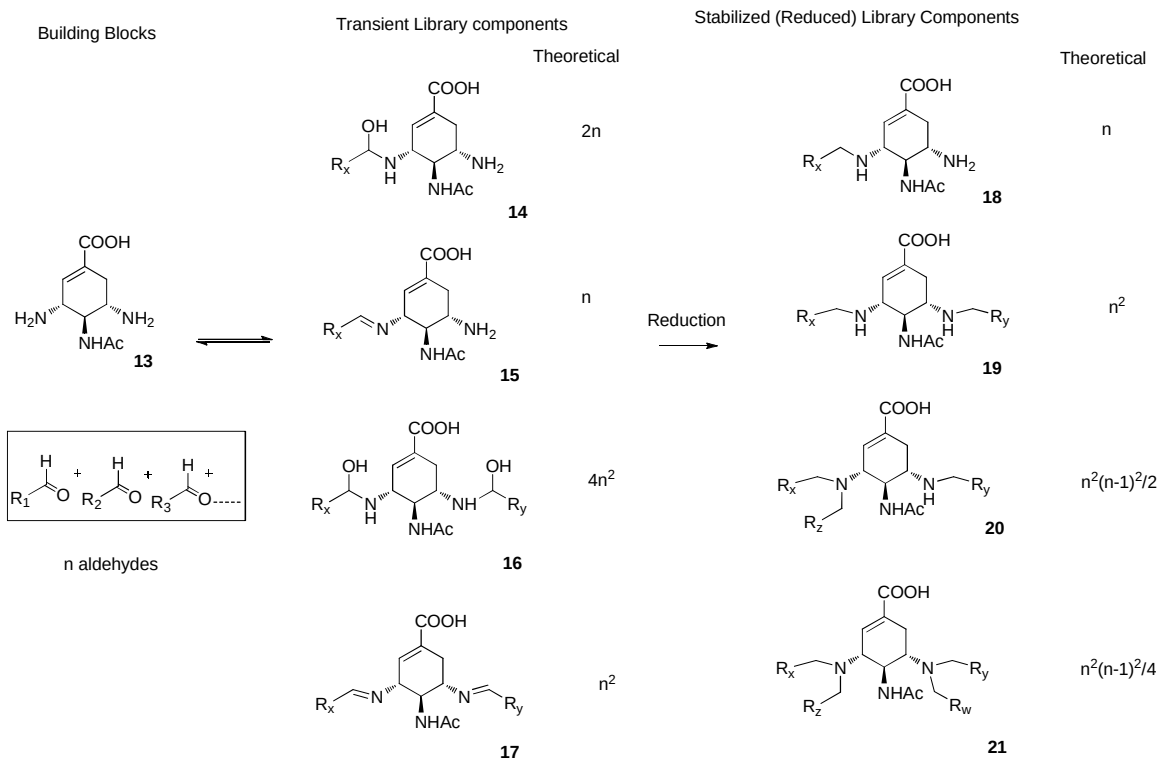
Large DCLs using simple imine chemistry targeting the enzyme neuraminidase were synthesized by Eliseev and co-workers.[13] Neuraminidase is an important drug target for influenza and is inhibited by molecules like tamiflu **12**.



Tamiflu **12**

Figure 1.6 *Anti-influenza drug Tamiflu*

A diamine compound **13**, structurally similar to tamiflu, was chosen as the scaffold for DCL generation. Aldehydes chosen from a pool of 41 commercially available compounds were added to the diamine **13** in the presence of sodium cyanoborohydride. The initial imine DCC products are reduced *in situ*, and analysed by HPLC as in the case of Lehn and Huc's pioneering work. This library with just 20 aldehydes can theoretically generate 40000 library members counting the hemi-aminals which have been previously seen by ^{1}NMR of sample DCLs. A unique feature of this DCL is its n x 1 stoichiometry: the use of a single, divalent amine as the scaffold which can, in principle, lead to the amplification of a single compound. Such amplifications are rare in DCLs with n x n (n > 1) stoichiometry.



Scheme 1.2 *Imine DCL generated from diamine* **13** *and n aldehydes*

A novel analytical method was adopted for the DCL. Given the large size of the library, Eliseev reasoned that it would be too difficult to attempt a detailed characterisation of all the compounds present at equilibrium. Instead, the DCL was generated under conditions where the transient compounds were present in extremely low concentrations and could not be detected by HPLC-MS in the absence of the target. However, in the presence of the enzyme, amplification of the best binders led to detectable amounts of the compounds which could be characterised by HPLC-MS. In this case compounds **18a, 18b** and **18c** were amplified with respect to the blank composition (Figure 1.7). Control experiments using bovine serum albumin instead of neuraminidase as the target protein produced no amplification which confirmed that the DCL evolution was controlled by the neuramindase active site. Regenerating the DCL in the presence of excess of zanamivir, a commercially available potent neuraminidase inhibitor, also gave a library composition similar to the blank. The absence of well-characterised blank libraries made it impossible to predict if the amplification observed was the result of a genuine thermodynamic selection or kinetic effects were involved. But given the precedence of imine chemistry used

before in DCLs it would not be unreasonable to assume that a DCL effect was operational in the system.

Biological assays of the re-synthesized hit compounds revealed that compounds **18a** and **18c** have increased activity relative to scaffold **13**. But, surprisingly amine **18b** which was amplified in the DCL is not a good binder exhibiting only a slightly higher activity over the scaffold. Amine **18d** on the other hand was not amplified at all in the DCLs but is a reasonably better inhibitor. These false positive and negative results are an inherent weakness of the imine DCL methodology and arise from the separation of the molecular recognition events that control DCL evolution and inhibitory activity assay. A subsequent report by Eliseev and co-workers extended the study of their DCC system to ketones instead of aldehydes.[14]

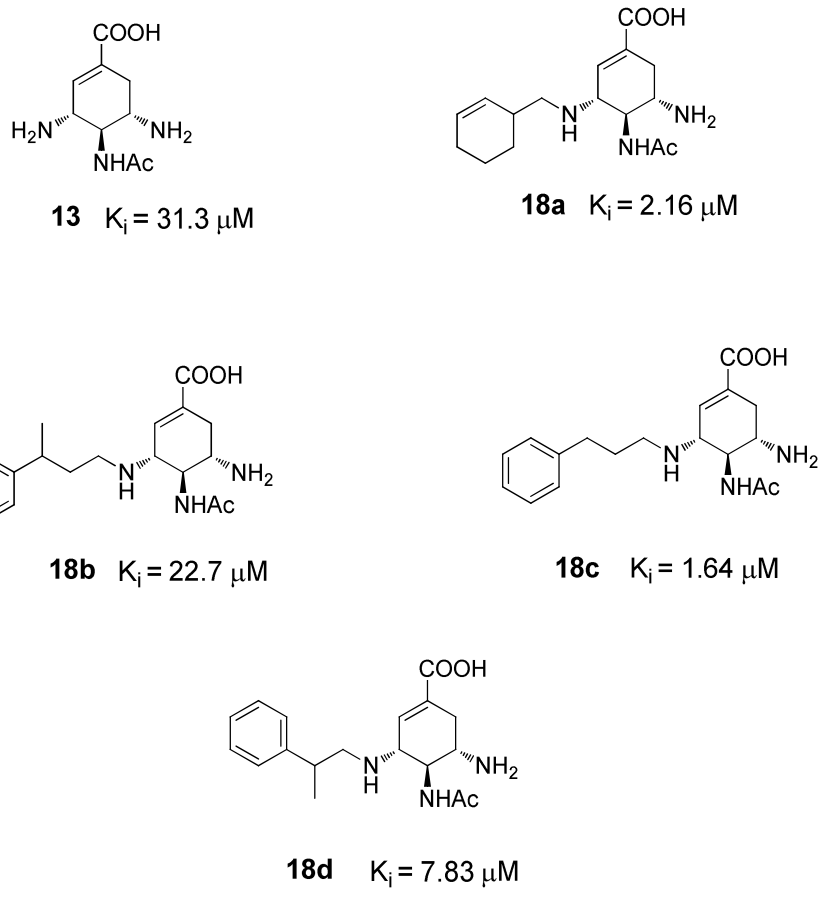
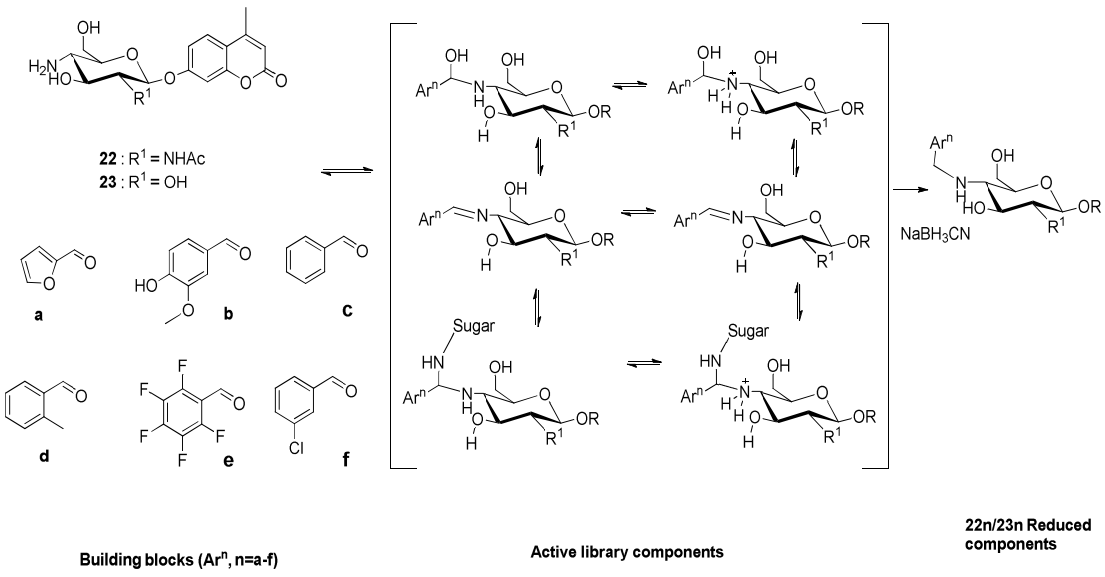


Figure 1.7 *K_i values for selected compounds from imine DCLs*

Beau has used DCC with a number of carbohydrate systems.[15-18] Carbohydrate-binding proteins are challenging targets since they often exhibit weak ligand interactions with millimolar dissociation constants. An example is hen egg-white lysozyme (HEWL), a glycosidase known to bind N-acetyl-D-glucosamine (D-GlcNAc) with low millimolar affinity. Choosing the imine bond as the reversible linkage to set up the DCL, the two D-GlcNAc and D-Glc compounds **22** and **23** along with equimolar amounts of six benzaldehydes (**a-f**) were equilibrated at pH 6.2 in the presence of sodium cyanoborohydride (Scheme 1.3). UV-HPLC analysis, made possible by the 4-methylumbelliferyl chromophore appended beforehand to the carbohydrate anomeric position, identified all the expected amines. The DCL equilibrated in the presence of stoichiometric amounts of HEWL led to a small amplification of two components, **22b** and **22f** (Figure 1.8). A control experiment with chitotriose, a good HEWL inhibitor, did not show any amplification effects. Biological assay confirmed that the amplified amine is about 100 times more active than the starting D-GlcNAc scaffold.



Scheme 1.3 *DCL of D-GlcNAc imines*

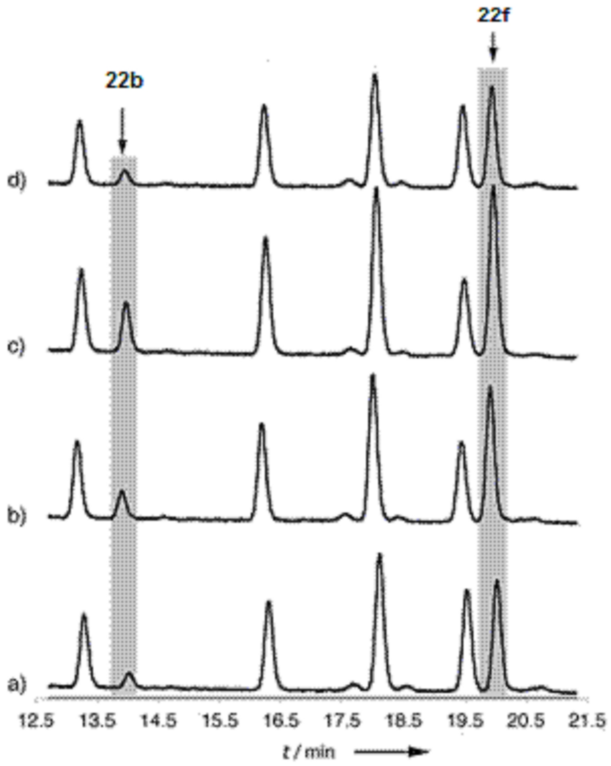


Figure 1.8 *Amplification studies of Beau's imine DCL targeting HWL. a) no HWL, b) 1 equivalent of HWL, c) 3 equivalents of HWL, d) 1 equivalent of HWL and 3 equivalents of chitotriose. Adapted from reference* 15.

Beau and co-workers then moved on to the more challenging glycosyltransferases (GTs) as targets.[16] The DCC approach is highly suitable in this case due to the absence of GT crystal structures. But a drawback lies in the requirement of stoichiometric amounts of the target protein necessary for measurable amplifications since GTs are typically available in very small amounts. This issue prompted Beau to formulate an interesting approach to imine DCLs. Assuming that the reduced amines share the binding characteristics of the parent imines and if a sub-stoichiometric amount of protein is used, competitive binding of the amine products to the target may decrease amplification over time. On the other hand if the amine products bound significantly more weakly than the imine DCL members, there would be no

competitive inhibition and constant amplification of the reduced amine products would be observed. A carbohydrate-based imine DCL (Figure 1.9) envisaging this scenario was designed targeting α-1,3- galactosyltransferase (α1,3GalT).

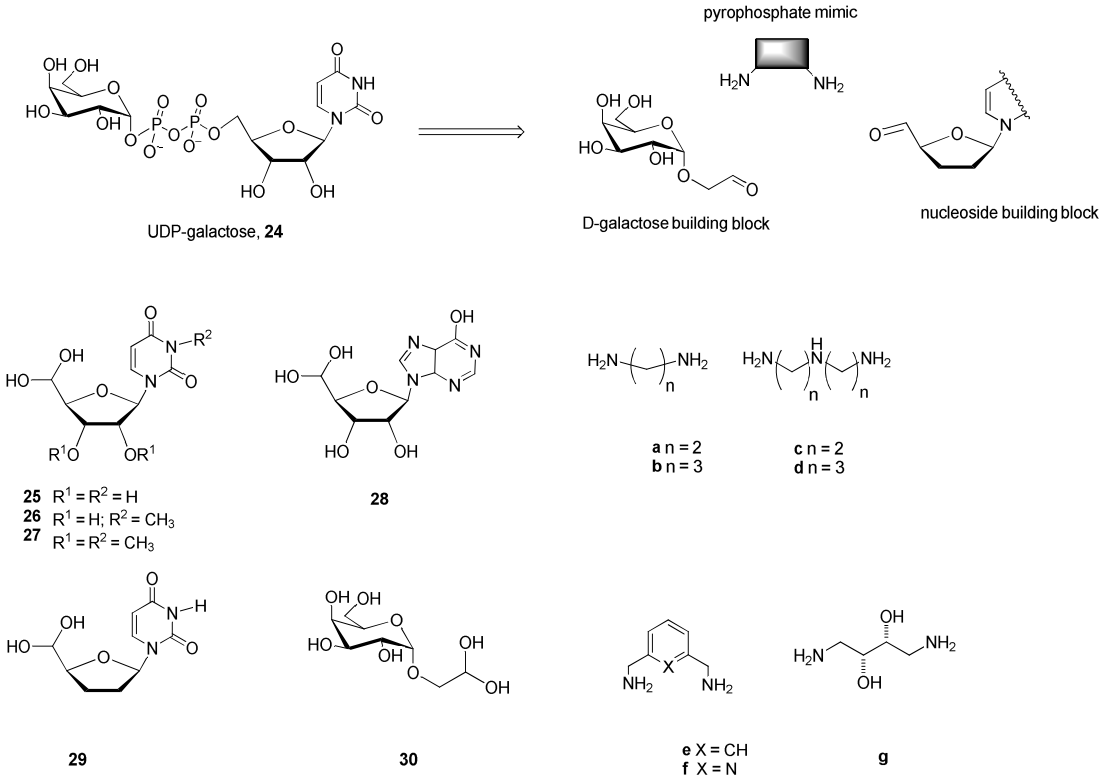
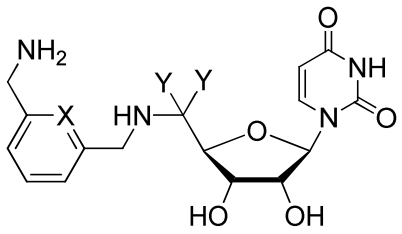


Figure 1.9 *Library building blocks for imine DCL templated by α1,3GalT*

The DCL had three types of components: Nucleoside aldehydes **25-29**, diamine linkers **a**–**g**, and D-galactose aldehyde **30**. The GT target concentration at 1.6 μM was significantly lower than the aldehyde and amine concentrations (82 and 542 μM respectively). Double amination, where both the nucleoside and galactoside aldehyde had condensed with a diamine, was not observed. The GT - templated DCL amplified monoamine products **25e** and **25f** relative to **25a, 25b** and **30f**. No amplification was observed in a control experiment with BSA, indicating molecular recognition by the GT target as being responsible for amplification. Inhibition assays established IC_{50} values of > 5mM for both amines (UDP-Gal IC50 = 73μM), with the

amides **25'e** and **25'f** showing values of 1.1 and 0.4 mM, respectively (Figure 1.10) which justifies the initial reasoning that $K_{amine} << K_{imine}$ would be necessary for amplification with the amides being 'geometric' mimics of the imine DCL components.

Extending their work on galactosyltransferases to the β-1,4-galactosyltransferase enzyme as a target for their carbohydrate DCL, Beau and co-workers observed only modest amplifications for amines **25a** and **25b**. This result is not very surprising since β-GT forms the glycosidic bond with inversion of configuration, whereas α-GT does so with retention.



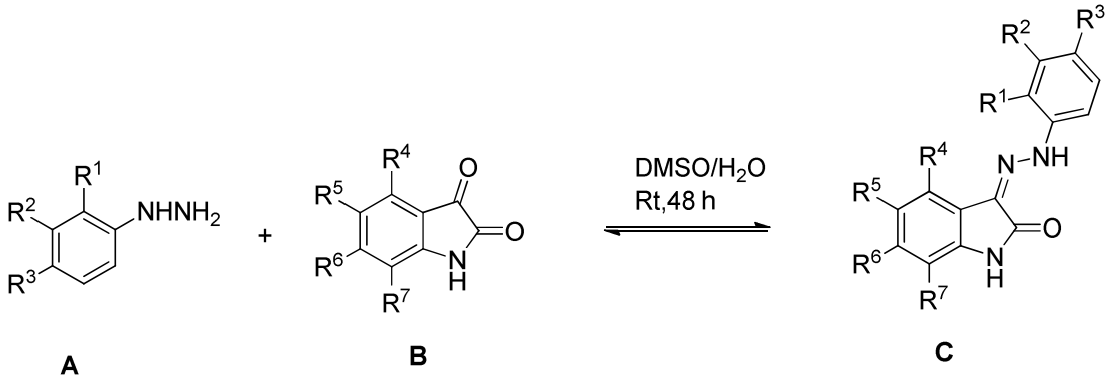
25e X = CH, Y = H IC_{50} = 5 mM
25f X = N, Y = H IC_{50} = 5 mM

25'e X = CH, Y = (=O) IC_{50} = 1.1 mM
25'f X = N, Y = (=O) IC_{50} = 0.4 mM

Figure 1.10 *Assay data for DCL hit compounds*

Congreve and co-workers introduced the concept of dynamic combinatorial X-ray crystallography (DCX) as a method for fragment-based drug discovery.[19,20] The technique has unique advantages over standard DCC approaches since along with direct identification of the ligand, the electron-density maps aid in understanding the detailed binding mode of the hit compound. The authors studied a kinase enzyme cyclin-dependant kinase 2 (CDK2) as the target with a DCL based on the aryl hydrazine **A** and isatin **B** scaffolds (Scheme 1.4). The library members were selected so as to present a range of functionality to the lipophilic pockets in the ATP binding

site of the protein. The suitability of hydrazone chemistry for the DCL experiment was established by reacting each of the 30 pairs of reactants in 20% aqueous DMSO over a period of 48 hours. After ensuring that all the expected hydrazones were being formed, the reactions were performed in the presence of CDK2 crystals first as individual pairs of hydrazine and isatin reactants, and then as mixtures. Analysing the electron density maps revealed a potent binder with inhibitory activity at 30 nM.



Scheme 1.4 *Hydrazone formation for dynamic combinatorial X-ray crystallography*

Although the DCX technique seems to be very powerful in ligand discovery applications, doubts remain as to whether a real DCC effect operates in the system since only a small amount of protein is used in the soaking experiments compared to the amount of library components in solution. The moot question is whether the protein structure influences the hydrazone equilibrium within the confines of the crystal or if a template effect operates under kinetic control.

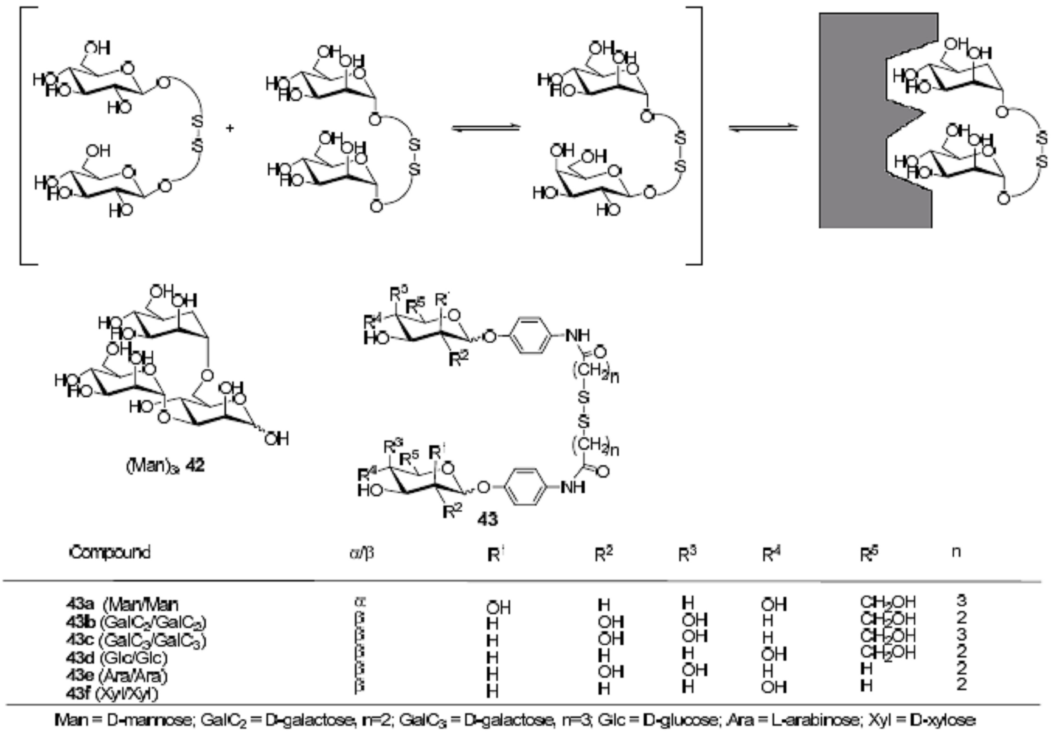
1.3.2 Disulfide bond formation

Disulfide bonds are ubiquitous in biology and play an important role in the folding of proteins and maintaining the redox states of cells.[21] The reversible nature of this bond has been known for some time and therefore it did not take very long for researchers to adapt it for DCC applications. The question whether a mixture of equilibrating disulfides can evolve and respond to template effects was answered for

the first time by Hioki and Still in 1998.[22] Their experiments used a DCL based on the disulfide exchange to identify a small molecule receptor for the tripeptide (D)Pro(L)Val(D)Val. After this initial report extensive work undertaken by the Lehn and Sanders groups in the late 1990s established the reaction as a powerful tool for DCC. [23,24]

Disulfide exchange proceeds readily in water under near neutral conditions (pH 7-9) which makes it suitable for use with biomolecular templates. The reaction can be operated in the presence of oxygen or catalytic amounts of thiol and a mild base. The exchange process is switched off by acidification which aids library analysis. These bonds are sufficiently stable under aqueous conditions and exhibit reasonable chemoselectivity tolerating a number of functional groups. The exchange process can be used with protein templates provided no thiols or disulfide groups are present on the surface of the protein. Any such groups embedded in the interior of proteins would not generally pose a problem since it would be difficult for small molecule thiols at low concentrations to access these groups.

Lehn and Ramström reported the first example of disulfide bond formation in a carbohydrate- based DCL targeting a protein.[23] Concanavalin A (Con A) is a well characterised carbohydrate- binding protein with high specificity for the branched trimannoside unit (Man)3 **42.** A DCL was designed based on an assumption that the two peripheral mannoside units in (Man)3 are responsible for binding interactions with the lectin while the central unit serves as a linker. Accordingly mannoside-based building blocks with a phenylamido chromophore and short chain alkyl thiols were synthesized. The thiols formed disulfides under appropriate conditions with the disulfide bonds functioning as reversible linkers between the two carbohydrate-based head groups. Since most non-aromatic disulfides have similar reactivity their equilibrium constants are close to unity. This feature ensured an iso-energetic DCL with all product structures expressed to a similar extent at equilibrium.



Compound	α/β	R^1	R^2	R^3	R^4	R^5	n
43a (Man/Man	α	OH	H	H	OH	CH_2OH	3
43b ($GalC_2$/$GalC_2$)	β	H	OH	OH	H	CH_2OH	2
43c ($GalC_3$/$GalC_3$)	β	H	OH	OH	H	CH_2OH	3
43d (Glc/Glc)	β	H	H	H	OH	CH_2OH	2
43e (Ara/Ara)	β	H	OH	OH	H	H	2
43f (Xyl/Xyl)	β	H	H	H	OH	H	2

Man = D-mannose; $GalC_2$ = D-galactose, n=2; $GalC_3$ = D-galactose, n=3; Glc = D-glucose; Ara = L-arabinose; Xyl = D-xylose

Scheme 1.5 *Disulfide exchange DCL templated by ConA*

The DCL based on the four disulfides **43a, 43d, 43e** and **43f** was equilibrated at pH 7.4 for 2 weeks with dithiothreitol (DTT) as an initiating agent and then analysed by HPLC (Figure 1.11). All the ten expected ditopic carbohydrate combinations could be identified. The Con A target immobilised on sepharose beads was then introduced into the DCL and equilibrated making it the first example of a solid-supported protein used in DCC. Product amplification was identified first by quantifying components that had been removed from the solution phase and then by quenching the DCL with acid and washing off bound components from the resin beads. There was a clear preference for the D-mannose homodimer followed by the D-mannose heterodimers. A slightly higher yield for the selected structures was observed if the receptor was present in the mixture from the beginning of the scrambling process as compared to it being added after pre-equilibration. HPLC resolution problems were encountered for larger libraries but even in such cases quenching and elution from

the beads clearly showed the expected mannose dimers bound to the immobilized protein.

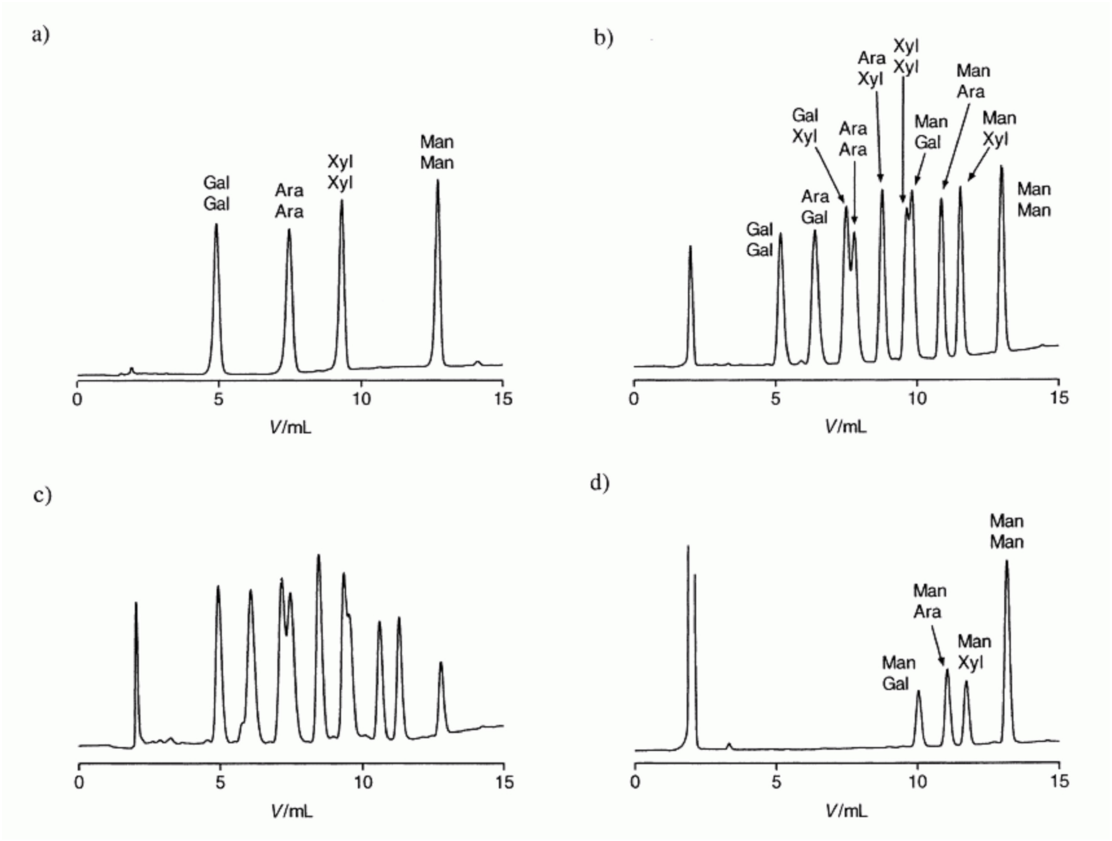


Figure 1.11 *LC analysis of disulfide DCLs. Trace a): Starting disulfides* ***43a, d, e, f****. Trace b): Equilibrated DCL (no target). Trace c): Equilibrated DCL in the presence of ConA. Trace d)Eluate of bound species from immobilised ConA. Reproduced from reference* 23.

Although the amplification effects in this study were relatively weak as expected for a low affinity system it clearly demonstrated the utility of mild disulfide exchange for DCC applications with protein targets. A similar DCL based on acyl hydrazone exchange targeting the same Con A binding site was subsequently reported by Lehn.[25] In this case due to the acidic conditions required by the exchange reaction a pre-equilibration / deconvolution approach was adopted.

The concept of disulfide-based DCLs targeting proteins was extended by Hunter and Waltho to the identification of bifunctional ligands for calmodulin (CaM).[26] This

protein target plays a major role in regulating a range of physiological processes by binding to enzymes. It has a unique structure with two independently mobile domains connected by a flexible linker. Therefore, the design of the DCL targeting CaM was based on a scaffold consisting of two hydrophobic peptides joined together by a flexible linker. The library components consisting of cystine dimers of hydrophobic amino acid residues are shown in Figure 1.12.

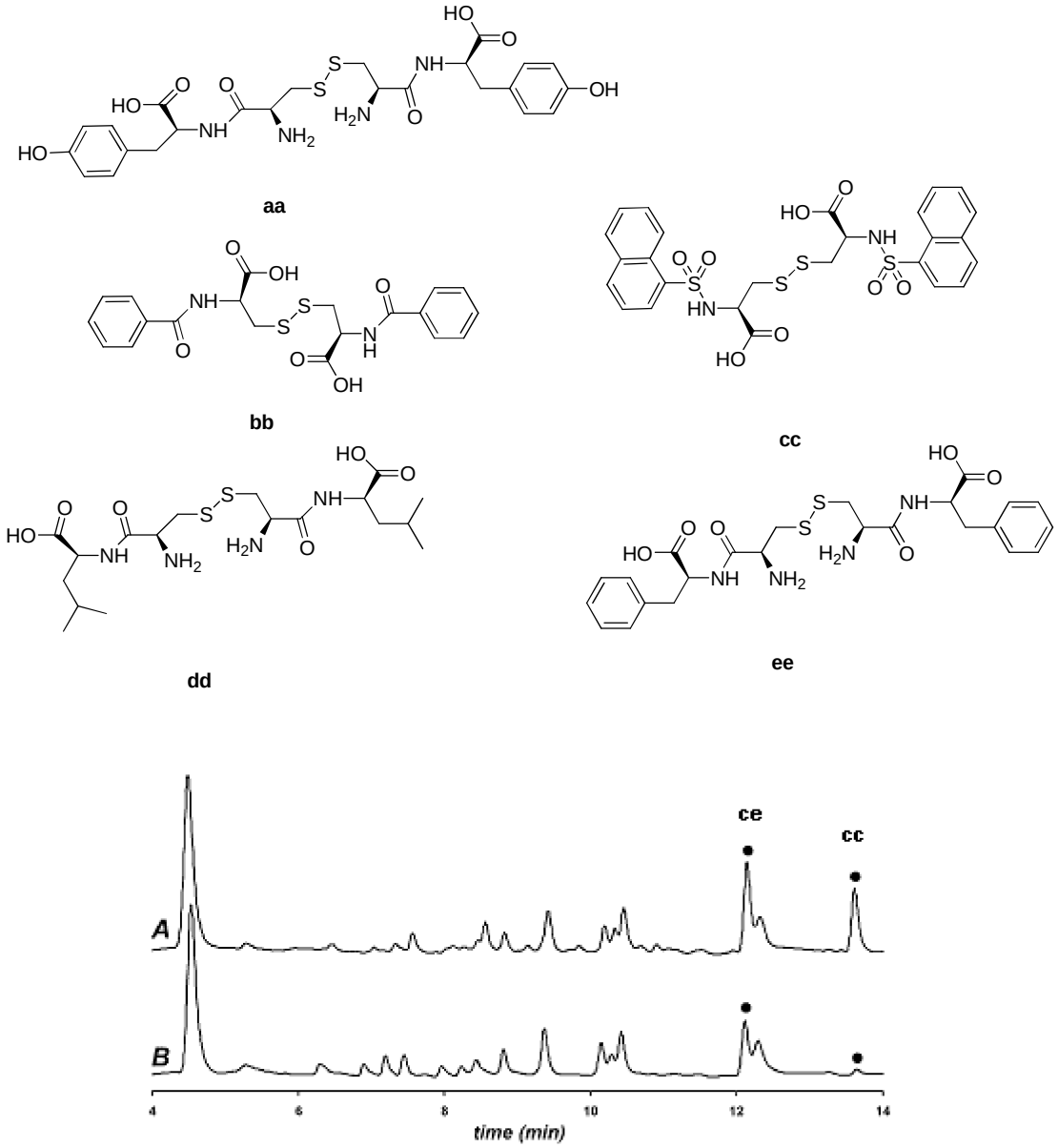


Figure 1.12 *Cystine dimer components for Calmodulin DCL and LC analysis of cystine DCL. Trace A shows the DCL templated with CaM while trace B shows the blank DCL. Dipeptides* ***ce*** *and* ***cc*** *are amplified (Reference* 26*)*

The DCL upon equilibration for 48 hours under standard disulfide exchange conditions at pH 7.5 expressed all the 15 expected disulfide exchange products as analysed by LC-MS (Figure 1.12). The library equilibrated in the presence of CaM was subjected to a centrifugation / filtration step to separate the protein / bound components from the mixture. Unfortunately the filtration membrane affected the library composition and HPLC analysis of the filtrate did not yield any meaningful results. However, analysing the bound components by denaturing the protein and comparing with the blank library composition indicated significant amplification for the dimer **ce** (80%) and a smaller amplification for the dimer **cc** (10%). A binding assay performed for these bifunctional ligands established K_d values of 10 and 210 μM respectively. These affinities were two orders of magnitude greater than the corresponding monofunctional ligands. This elegant study proved the use of disulfide chemistry in DCLs targeting more complex proteins.

Danieli and co-workers adopted a similar approach to design a DCL of bivalent ligands linked by flexible disulfide linkers targeting the tubulins.[27] The antitumour agents thiocolchicine and podophyllotoxin were chosen as the scaffolds for the library. These compounds were appropriately derivatised to form homo and hetero-dimers through disulfide linkers (Figure 1.13). An *in situ* templation with the intended target microtubulin did not work due to the poor aqueous solubility of the DCL members. The tubulin itself collapsed in organic solvents which ruled out its use in an acetone –based DCL. Albumin and subtilisin, two organic solvent tolerant proteins used as alternative targets, influenced the amounts of the homodimers **45** and **48** and the hetero-dimer **51** in a DCL generated in acetone and triethylamine (TEA).

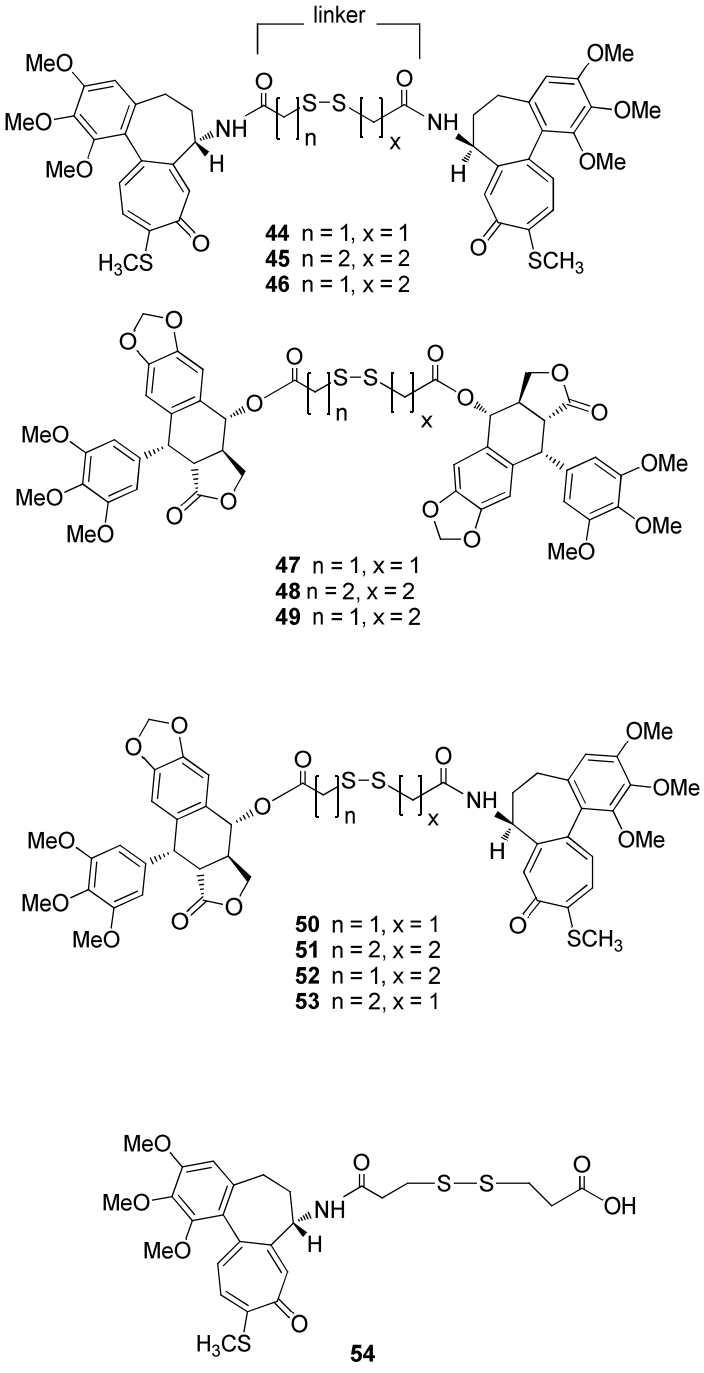
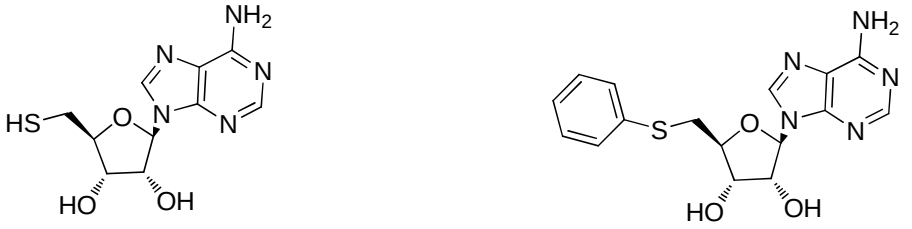


Figure 1.13 *Disulfide DCL based on thiocolchicine and podophyllotoxin*

Abell and Ciulli have recently combined ideas in fragment based drug discovery with DCC to probe adenosine binding sites on the enzyme *Mycobacterium tuberculosis* pantothenate synthetase.[28] They used 5'-deoxy-5'-thioadenosine (**55**) as an anchor building block which can undergo thiol-disulfide exchange templated by the enzyme.

Library analysis by HPLC led to the identification of a benzyl disulfide derivative (**56**) which was susequently derivatised to generate novel compounds that bind strongly to the pantoate site of the protein.



55 **56**

Figure 1.14 *Anchor building block (**55**) targeting the pantoate site and the amplified benzyl disulfide derivative(**56**)*

Erlanson *et al* working at Sunesis Pharmaceuticals tried to harness the power of disulfide exchange chemistry in a conceptually new fragment-based drug discovery method termed tethering.[29-35] This approach has many features common with established DCC protocols and is illustrated in Figure 1.15. It uses a cysteine residue on the target protein surface which undergoes reversible disulfide exchange to select small molecule thiol fragments from solution. These weak-binding thiol fragments can then be elaborated to develop drug candidates with better affinity for the target protein. A limitation of this methodology is the requirement of a cysteine residue close to the active site of interest. If the protein does not have the required cysteine residue at a suitable location then they have to be introduced using site-directed mutagenesis. Though this is a well established procedure in protein chemistry it still requires structural information about the protein to exactly pinpoint the location for the thiol to be introduced on the protein surface.

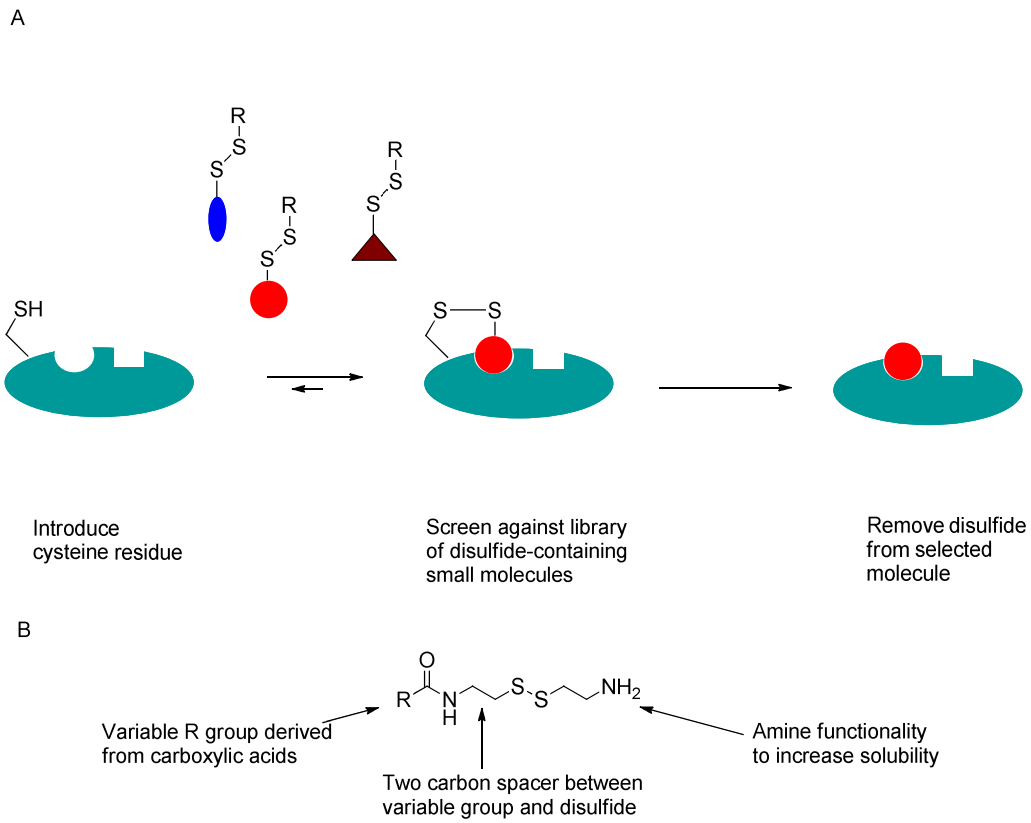


Figure 1.15 *The concept of tethering*

The first successful demonstration of this approach used the enzyme thymidylate synthase (TS) as the target.[35] This protein is a well characterised anti-cancer drug target with an active site cysteine. The small molecule library was designed around the general scaffold structure (Figure 1.15). It contains a variable fragment part derived from a carboxylic acid that could be connected to a terminal amine through a disulfide linker. The role of the terminal amine is primarily to increase solubility of the components in aqueous solution. Library sizes were in the range of 8-15 disulfides with individual concentrations at 0.2 mM. The disulfide bonds were exchanged under standard conditions in the presence of 15 μM protein followed by LC-MS analysis where the hit compounds covalently bound to the proteins were identified. The MS analysis assumed that the ionisation profile of each tethered protein-complex is similar with each tethering experiment featuring fragments of distinct molecular mass differing by at least 10 Da.

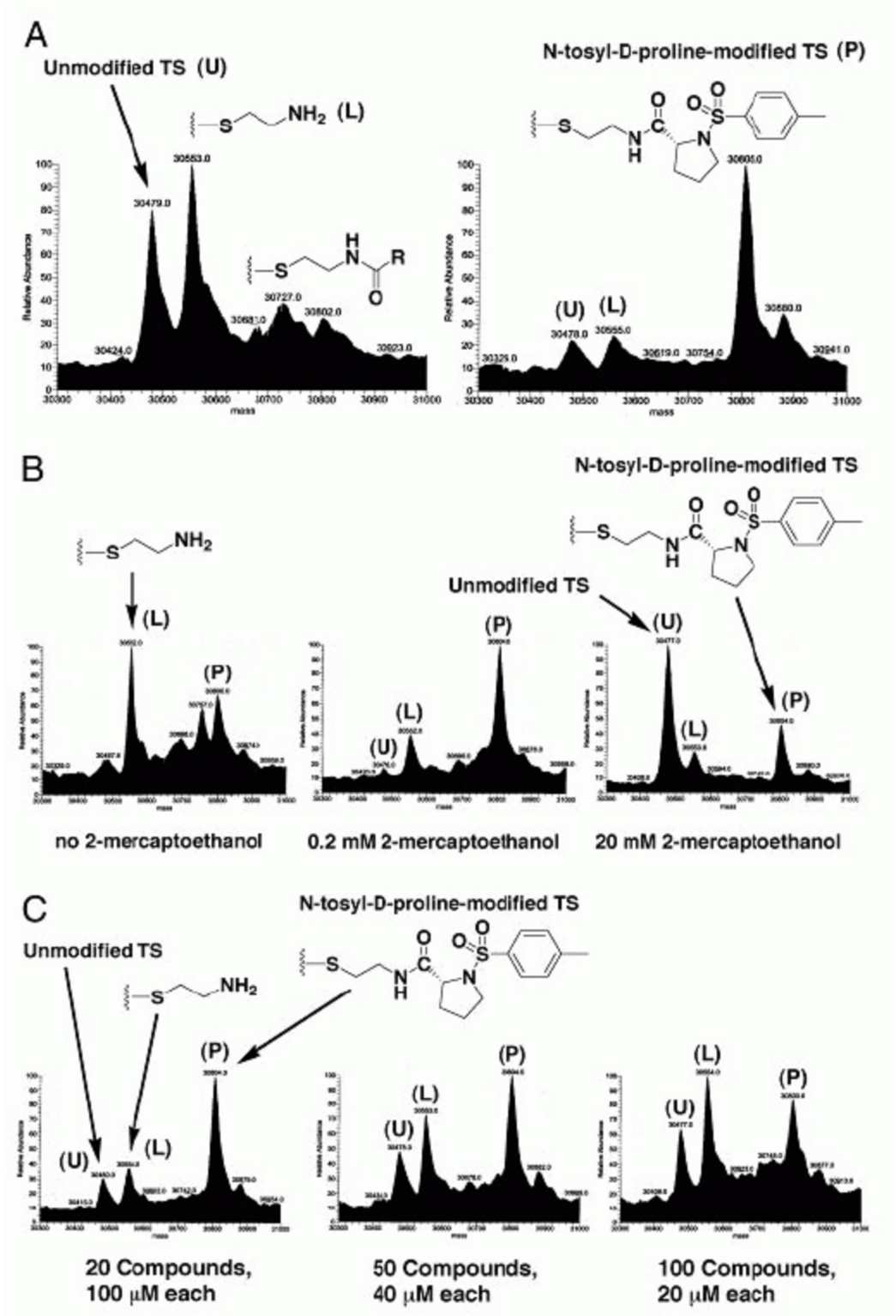


Figure 1.16 *MS data of tethered complexes. Data A shows a tethering experiment between TS and 10 disulfides equilibrated for 1 hour. Data B illustrates tethering with varying concentrations of 2-mercaptoethanol. Data C shows tethering with varying pool size of disulfides. Reproduced from reference* 35.

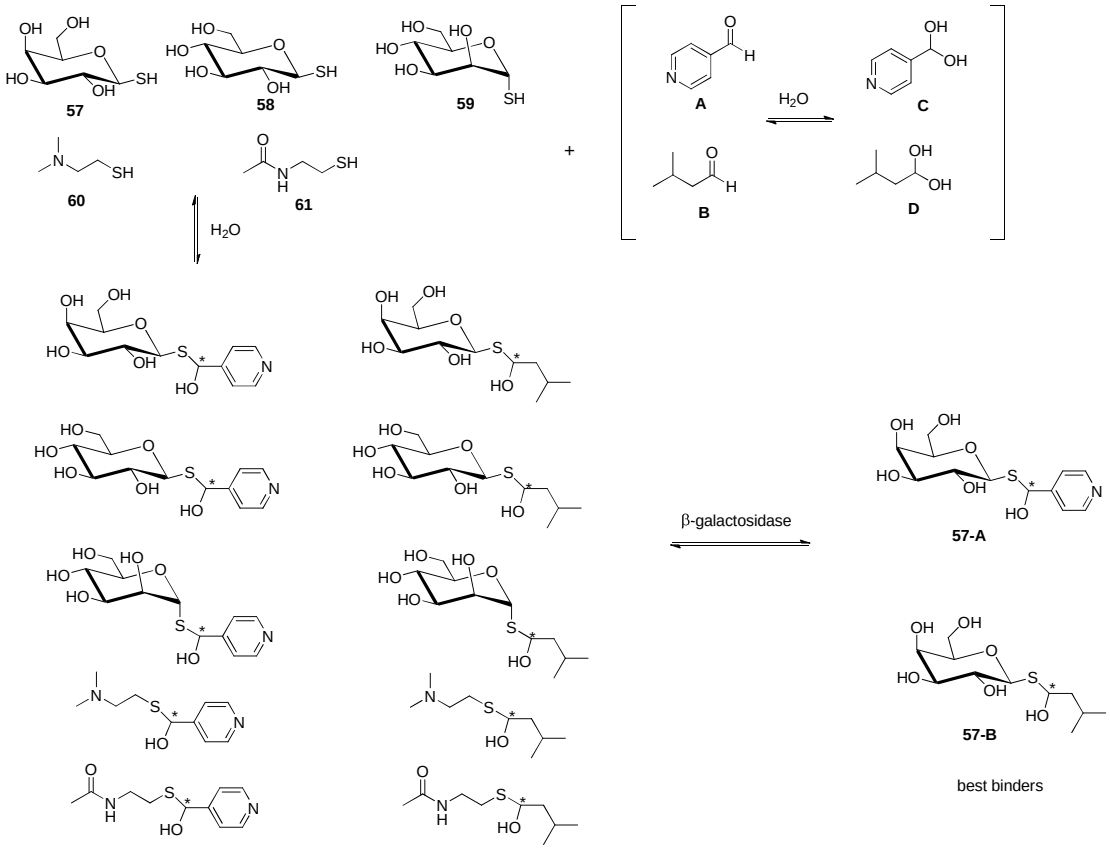
Figure 1.16 shows the results of these tethering experiments. For experiments with no hits the deconvoluted MS shows peaks for the unmodified protein (U) along with TS tethered to 2-aminoethanethiol (L) which is common to all the library components. A clear hit was observed with an N-tosyl-D-proline fragment (Figure 1.16A). Thermodynamic control over the selection was proved by varying the amount of catalytic thiol for disulfide equilibration (Figure 1.16B). An advantage of the tethering technique is its scalability with the N-tosyl-D-proline fragment chosen from experiments containing over 100 compounds (Figure 1.16C). The study was extended to a library of 1200 compounds based on the proline fragment P thus producing a SAR picture which was supplemented with crystal structures of the protein with the fragment bound at the active site. Significantly the conformation of fragment P in the tethered complex was very similar to a non-covalent complex between N-tosyl-D-proline and TS which shows that the covalent linkage in the tethering experiment does not affect the binding mode of the fragment. Inhibitors of vastly improved affinity were then developed by linking up the side chain of the TS cofactor methylenetetrahydrofolate. Based on this initial success the tethering concept has since then been applied to a number of targets.[30-34]

1.3.3 Hemithioacetal formation

Recently Ramström *et al* have reported the use of hemithioacetal formation for DCC applications.[36] Addition of thiols to aldehydes and ketones leads to the reversible formation of hemithioacetal stereoisomers but the equilibrium lies towards the starting components. This results in virtual dynamic systems with only transient amounts of the hemithioacetals detected in solution. The library used by the authors consisted of five thiols and two aldehydes generating 10 transient hemithioacetals (20 stereoisomers) in aqueous conditions (pH 7.2, ambient temperature) suitable for a protein (Scheme 1.7). The target protein, β-galactosidase, was chosen on the basis of the absence of cysteine residues at the active site which would prevent any competing hemithioacetal formation with the free aldehyde library components. The protein is a well-characterised hydrolase catalysing the hydrolysis of glycosidic linkages of β-galactosides into the corresponding galactoses and alcohols.

Under the conditions of library generation (pH 7.2, ambient temperature) in the absence of the enzyme no characteristic NMR signals were seen for hemithioacetals but upon reducing the pH from 7.2 to 4 specific hemithioacetal signals were observed slightly downfield from the hydrate signals of the corresponding aldehydes.

For the enzyme-templated virtual DCLs Ramstrom and co-workers used saturation transfer difference (STD) NMR spectroscopy to identify bound ligands. In this technique magnetization is transferred from protein active site protons to those of the bound ligand which is detected upon ligand dissociation provided the relaxation time scale is slower than the rate of dissociation. Hemithioacetals **57-A** and **57-B** were identified as the best binding transient species which was later confirmed by inhibition studies.



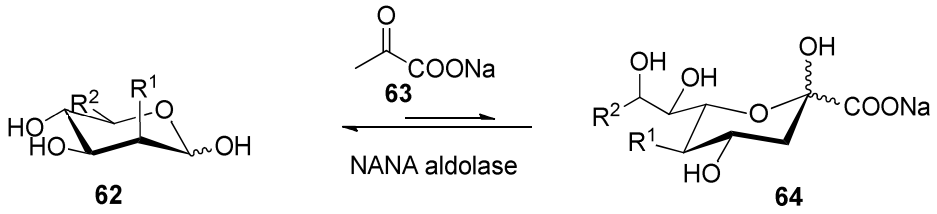
Scheme 1.7 *Virtual dynamic hemithioacetal system targeting β-galactosidase*

1.3.4 Enzymatic methods

An early example of DCC using enzyme-catalysed reactions was reported by Venton *et al* in 1997. [37] Catalysing reversible reactions using enzymes is a practically sound idea since they work under physiological conditions and can be applied to a range of C-C and C-X bond forming reactions. Venton chose the protease-catalysed amide bond synthesis / hydrolysis for generating his DCL. For the protease-based DCL two small peptides YGG and FL were used as the starting points. Incubating these with the broad spectrum protease thermolysin yielded over 15 short peptide sequences. In another experiment incubating the dipeptides VA and AL with thermolysin followed by cathepsin C produced 7 out of the possible 9 dipeptides. These results proved that the protease scrambles the amino acid sequence of the starting peptides. To test the protease DCL with a suitable receptor, a mixture of BSA hydrolysates and the tripeptide GPR was incubated with thermolysin and fibrinogen separated by a dialysis membrane. Fibrinogen is a readily available protein involved in the blood clotting process while the tripeptide GPR is a known binding motif for the fibrinogen target. Analysing the composition of the templated mixture and comparing it with control experiments identified amplification of the peptides GPRL, GPRF and DKPDNF by fibrinogen. The peptides assayed for binding with fibrinogen gave K_a values in the range of 10^{-4} M^{-1}, an affinity two-fold lower than the tripeptide GPR. This study by Venton is impressive for its use of enzyme- catalysed amide exchange in a selection process. However reversible amide bond formation using chemical methods in a DCC context is still an open challenge though some progress has been reported recently by Stahl.[38-40]

A second example of an enzyme-catalysed reversible reaction used to generate a DCL was reported by Flitsch and Turner (Scheme 1.8).[41,42] The library was based on an aldolase enzyme-catalysed breaking up of sialic acid **64a** to N-acetylmannosamine **62a** and sodium pyruvate **63**. A small DCL consisting of N-acetylmannosamine **62a**, D-mannose and D-lyxose was set up in the presence of N-acetylneuraminic acid (NANA) aldolase. Equilibration of the aldol products took 16 hours while control experiments proved the system to be under thermodynamic control. Wheat germ agglutinin (WGA), a plant lectin and a known weak binder of

sialic acid was selected as the target of the DCL. A clear amplification was observed for the natural substrate sialic acid **64a** at the cost of the adduct **64b**. Thermodynamic measurements established the binding affinity of **64a** to WGA as 172 M^{-1}. No binding could be detected for compound **64b.**



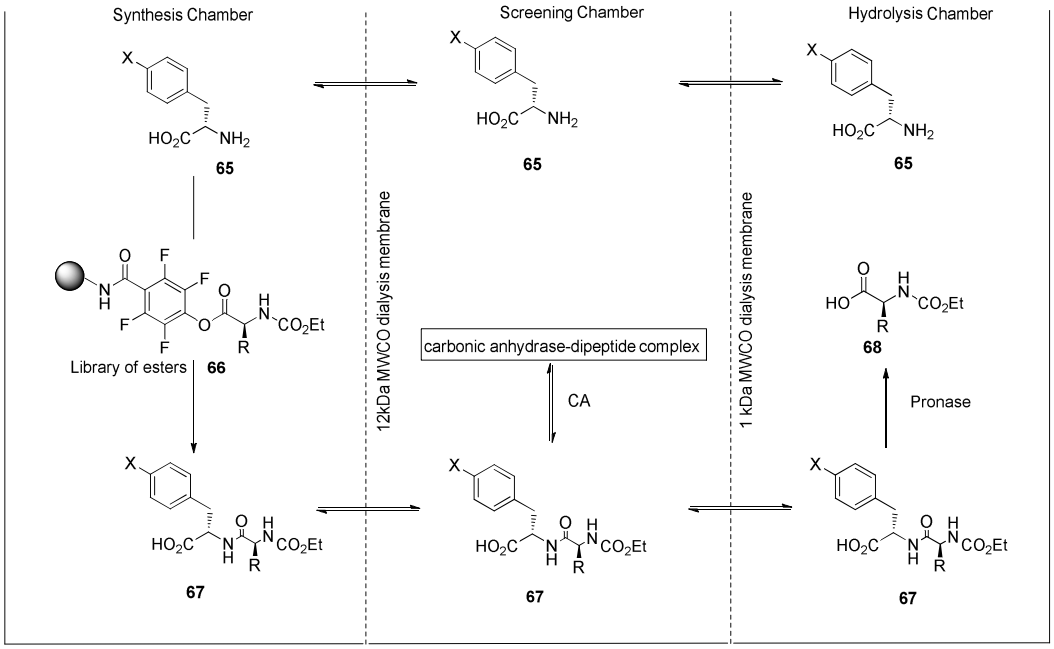
62a R^1 = NHAc, R^2 = CH_2OH
62b R^1 = OH, R^2 = CH_2OH
62c R^1 = OH, R^2 = H

64a R^1 = NHAc, R^2 = CH_2OH
64b R^1 = OH, R^2 = CH_2OH
64c R^1 = OH, R^2 = H

Scheme 1.8 *Aldol reaction catalysed by N-acetylneuraminic acid(NANA) aldolase to produce sialic acid*

1.3.5 Pseudodynamic combinatorial chemistry

Pseudodynamic combinatorial chemistry is a very elegant strategy introduced by Gleason and Kazlauskas in 2002.[43,44] They combined an irreversible synthesis of the library members with an irreversible destruction step. The best binding compounds form a complex with the target protein and survive the destruction. Scheme 1.9 illustrates the idea for a peptide-based pseudoDCL targeting carbonic anhydrase (CA). The reaction system is compartmentalised in a three-chambered reaction vessel; the first chamber functioning as the synthesis chamber, the second containing the target enzyme and thus functioning as the screening chamber, and the third functioning as the destruction chamber in which the weak binders are destroyed. Practically this is achieved by using dialysis bags suspended in a solution. Diffusion of components across the three chambers imparts a 'pseudodynamic' nature to the whole system while the irreversible synthesis and destruction steps vastly improve the selectivity of the process.



Scheme 1.9 *The pseudoDCL experiment*

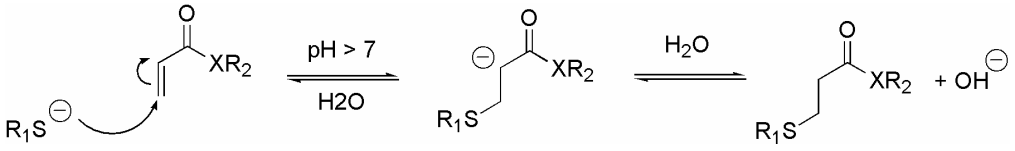
Dipeptide binders are created in the synthesis chamber through an addition of amino acids **65** to the immobilised ester **66**. These dipeptides (**67**) diffuse into the screening chamber and establish a binding equilibrium with the target CA. The free dipeptides which do not form a complex with the CA diffuse into the third destruction chamber where they are hydrolysed back to the amino acids using excess of a broad spectrum protease, pronase. This excess was used to ensure that the natural selectivity of the protease would not create large differences in the rates of the synthesis and destruction steps. The amino acid building blocks then migrate to the first synthesis chamber and are reused in the synthesis step. The stock of the ester **66** was constantly replenished. The higher affinity members which are protected from destruction are amplified at the expense of the poor binding molecules.

The designed pseudo-DCL used two amino acid building blocks along with four immobilised esters. After seven cycles of replenishing the esters (112 hours) the best binding dipeptide was amplified by a factor of 100 over the next best binder even though the ratio of their binding constants is only 2.3:1. This increased selectivity of

the pseudodynamic approach in comparison to classical DCLs may be due to the iterative nature of the process.

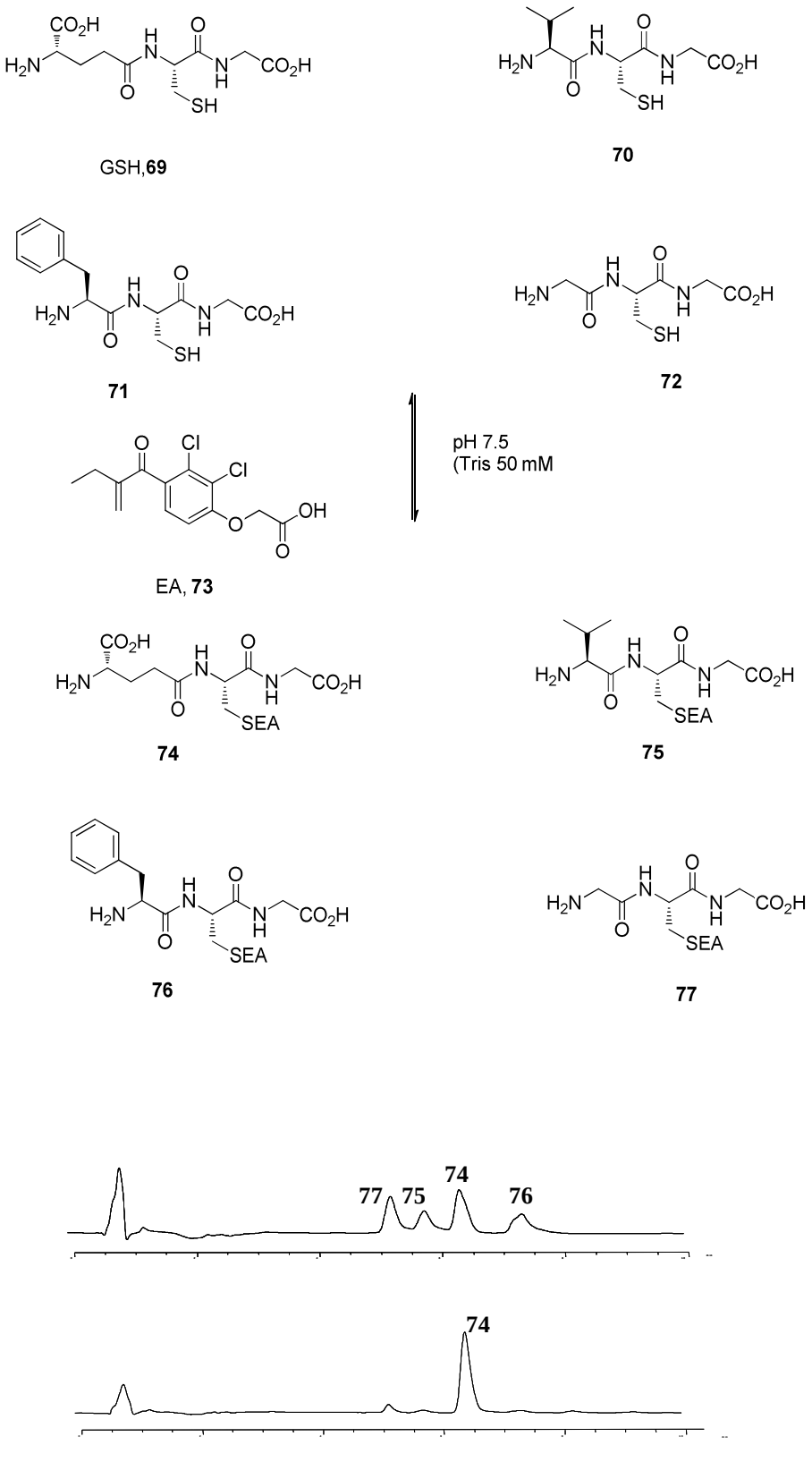
1.3.6 Michael addition of thiols to enones

Recent work by Greaney and co-workers has led to the introduction of the conjugate addition of thiols to Michael acceptors as a new strategy for generating adaptive DCLs (Scheme 1.10).[45,46] The reaction takes place in water under mildly basic conditions at room temperature which permits its use with biological templates. This exchange process is reasonably fast at physiological conditions and extremely slow and irreversible at low pH, allowing the reaction to be controlled by a simple change in pH.



Scheme 1.10 *Reversible Michael addition of thiols to enones*

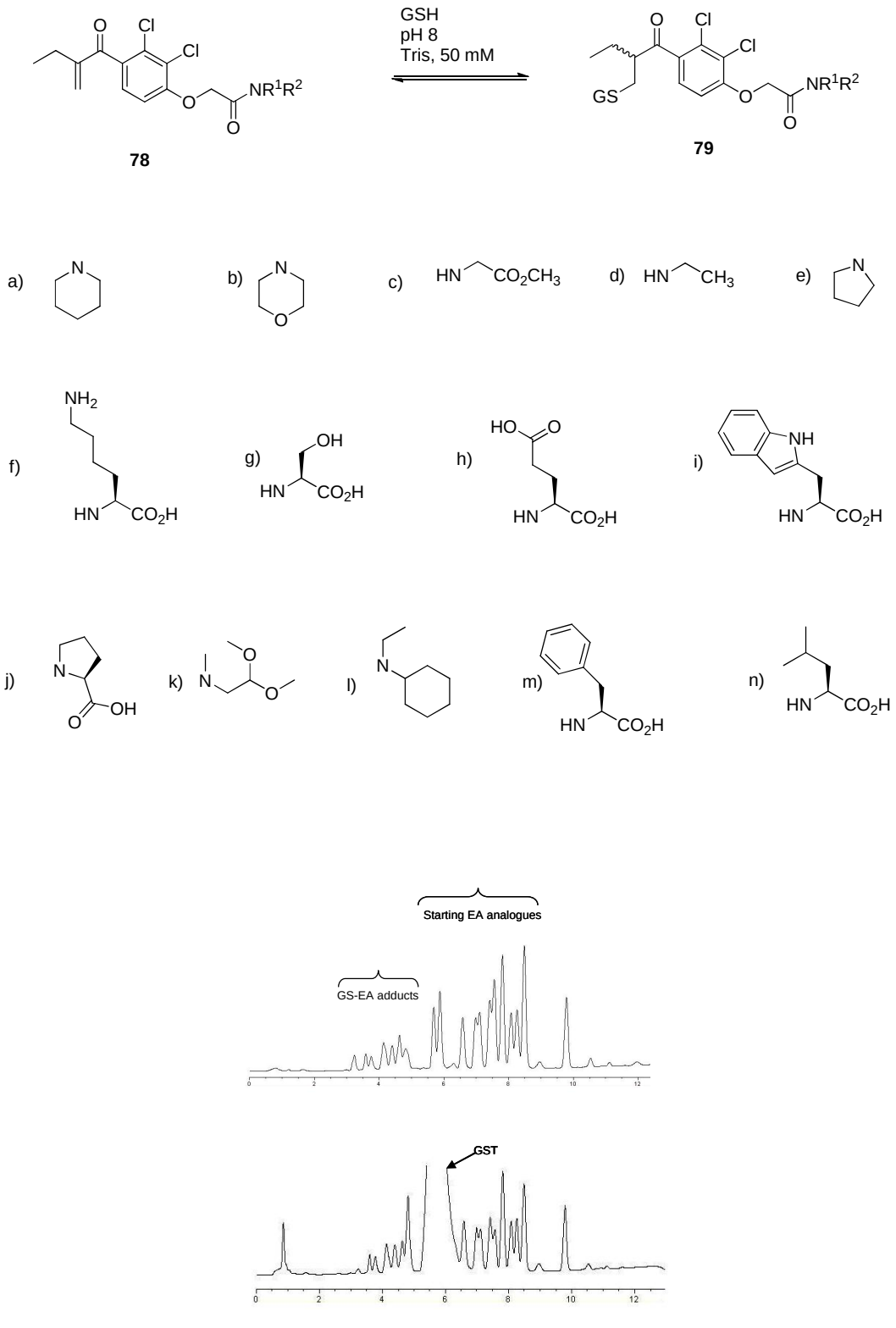
The enzyme glutathione-S-transferase (GST) from the helminth worm *Schistosoma japonicum* was selected as the target for the conjugate addition DCC experiments. These enzymes catalyse the conjugation of glutathione (GSH, **69**) to a wide range of electrophilic substrates thereby aiding cell detoxification and are therefore considered to be important drug targets for cancer and parasitic diseases like schistosomiasis and malaria.



Scheme 1.11 *Thiol conjugate addition DCL templated by SjGST (Reference 46)*

A Michael acceptor, ethacrynic acid (EA, **73**), GSH (**69**) and three tripeptide analogues (**70-72**) were used to generate a preliminary DCL probing the GSH binding site of SjGST. Ethacrynic acid **73** is a known inhibitor of the GST class of enzymes and binds at the hydrophobic region of the active site. The four expected Michael adducts **74-77** equilibrated in an hour and were identified by LC-MS. Upon addition of an equivalent of enzyme, the native substrate GS-EA (**74**) was amplified to 92% of total adducts concentration compared to 32% in the absence of the protein. Binding assays performed on the compounds showed a difference of an order of magnitude between the selected compound **74** (IC_{50} = 0.32 μM) and compound **77** (88 μM)

A larger library consisting of a single thiol, GSH, and 14 EA analogues was then set up to probe the hydrophobic binding site of the same protein (Scheme 1.12). This region of the active site is not conserved across the different GST isoforms and therefore would be crucial in the development of isoform-specific inhibitors. The templated library was found to amplify three compounds **78a, 78m** and **78n** chiefly at the expense of the lysine derivative of EA **78f**. Binding assays established that the selected compounds were more potent binders (IC_{50} = 0.61 μM for **78a**) than the lysine derivative **78f** (IC_{50} = 8.20 μM).



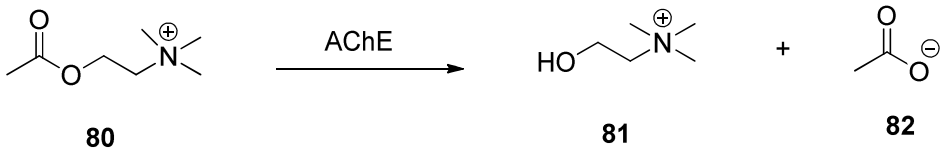
Scheme 1.12 *Thiol conjugate addition DCL templated by SjGST (Reference* 46*)*

1.4 Dynamic Combinatorial Resolution

Ramström introduced the concept of dynamic combinatorial resolution as a novel method for the kinetic screening of dynamic libraries.[47-49] The approach tries to resolve some of the difficulties associated with classical DCC: the analysis of complex mixtures and the requirement of stoichiometric amounts of the target molecule. A complete resolution of dynamic libraries in a one-pot process can be achieved using catalytic amounts of the target if the DCL and the binding event can be coupled to a second kinetically controlled step mediated by the target enzyme. This irreversible step may result in a stable product which is easy to isolate and identify.

1.4.1 Transthiolesterification DCR

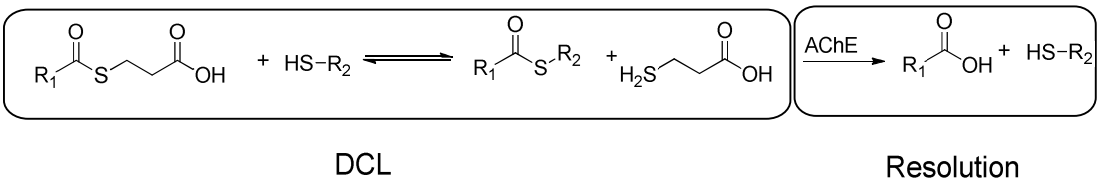
Initial work by Ramström and coworkers demonstrated the use of transthiolesterification for the generation of DCLs in neutral aqueous solutions, conditions highly desirable for biological targets.[47] These DCls were then subjected to a DCR process using hydrolases as targets. The target hydrolase of choice, acetylcholinesterase, is a serine hydrolase which plays a very important biological role in the hydrolysis of the neurotransmitter acetylcholine (**80**) to acetate and choline (**81**) (Scheme 1.13). The enzyme is highly efficient with hydrolysis rates approaching diffusion-controlled limits around 10^9 $M^{-1}s^{-1}$, a factor important in the design of a proof-of-concept DCR.



80 **81** **82**

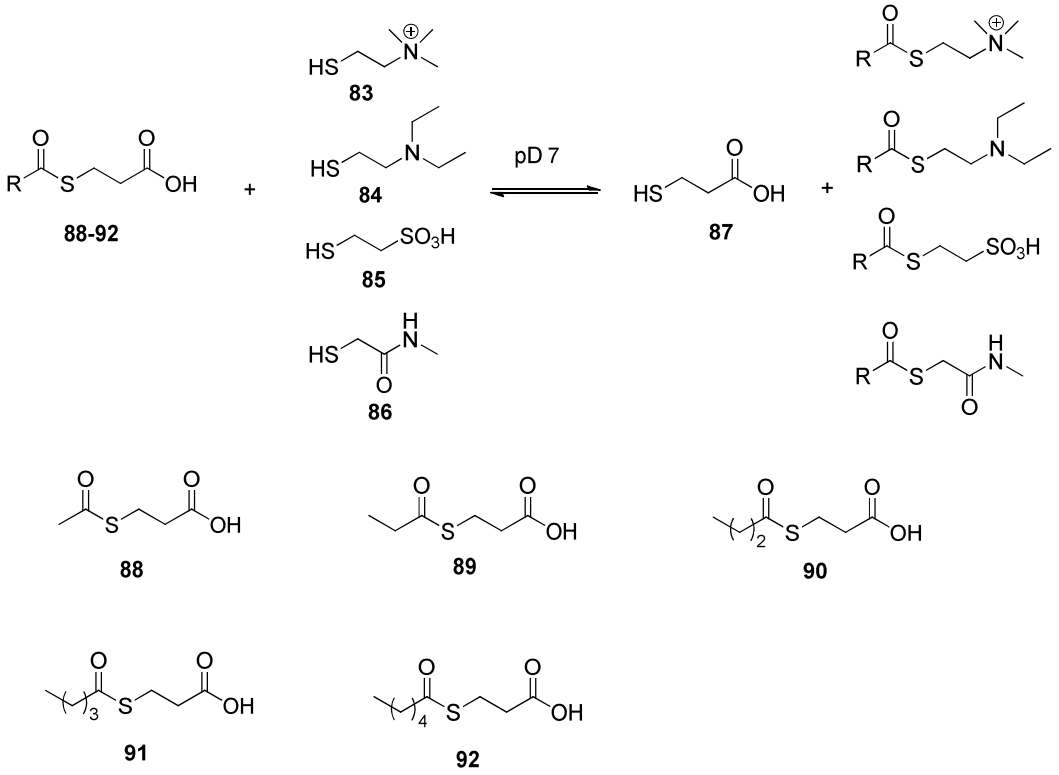
Scheme 1.13 *ACE catalyses the hydrolysis of acetylcholine*

The dynamics of the transthiolesterification depends on the structure of the thiols. Under neutral pH conditions the exchange rate was directly correlated to the pK_as of the thiols; those with pK_a less than 8.5 equilibrated rapidly. Thiolesters based on secondary thiols were considerably less stable than acetylcholine and the equilibrium favoured the reactants in the case of aromatic thiols. The rate of the exchange reaction can be increased by increasing the pH of the solution but this would also increase the rate of unwanted hydrolysis. The DCR concept using transthiolesterification is shown in Scheme 1.14.



Scheme 1.14 *Transthiolesterification DCR*

A 25-compound isoenergetic DCL was generated from a series of thiolesters and thiols followed by the addition of catalytic amounts of acetylcholinesterase. The protein immediately recognised the best substrates which were hydrolysed resulting in a loss of the acyl component from the library. The free thiol generated was accommodated in the dynamic library leading to an increase in the population of the hydrolysed species. The enzyme showed a clear preference for the acetyl and the propionyl groups with the acetyl species hydrolysed faster than the propionyl compound. After substantial hydrolysis of these two substrates the butyrate ester was hydrolysed at a smaller rate while the rest of the acyl groups remained untouched. A series of hydrolases tested with the library showed some activity depending on the respective enzyme specificities. Control experiments with bovine serum albumin (BSA) did not show any activity.



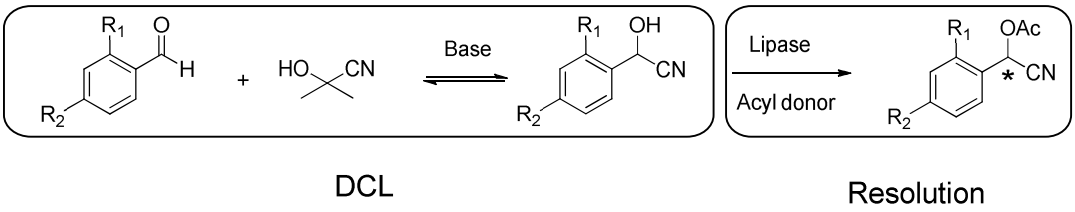
Scheme 1.15 *DCl for screening hydrolase substrates*

1.4.2 Cyanohydrin DCR

The DCR concept was extended by Ramstom *et al* to the screening of lipase substrates using cyanohydrin formation for the generation of the DCL (Scheme 1.16).[48] The reaction is an example of reversible C-C bond formation with the product cyanohydrins amenable to further functional group transformations. The lipase target enzymes are esterases which recognize a broad range of unnatural substrates to catalyse esterification reactions. The chief biological role of these proteins is to catalyse the hydrolysis of triglycerides into fatty acids and glycerol at water –oil interfaces.

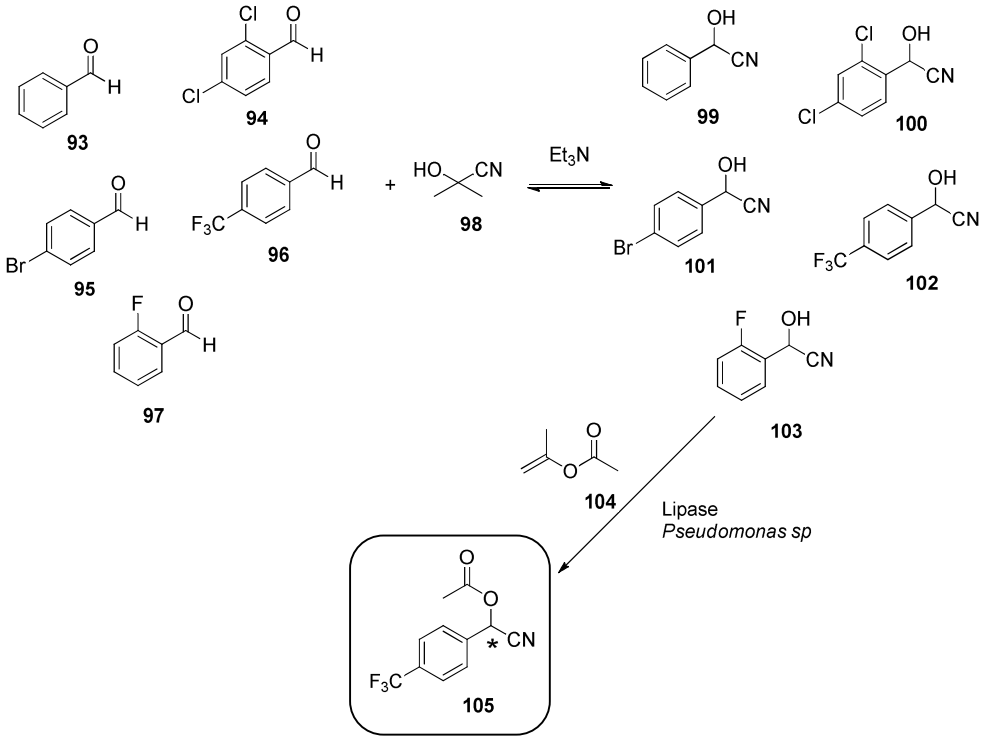
A unique method using acetone cyanohydrin adducts was used to generate cyanohydrin DCLs under mild conditions. Treating the acetone cyanohydrin with triethylamine base in chloroform releases the cyanide ion along with free acetone in

organic solvents. The released cyanide ions react with aldehydes or ketones to generate a dynamic library of stable cyanohydrin adducts.



Scheme 1.16 *Cyanohydrin DCR*

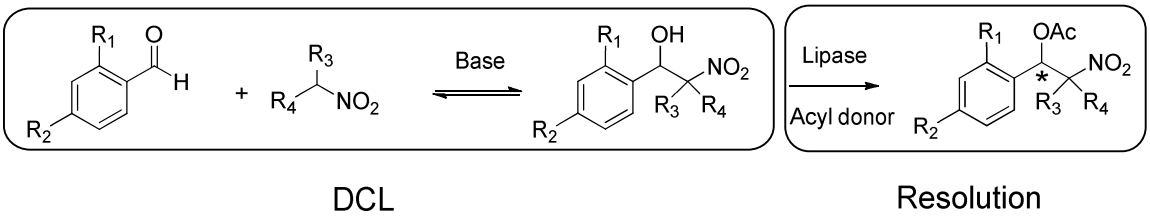
The authors used a small library of five different benzaldehydes along with acetone cyanohydrin and triethylamine which attained equilibrium in 3 hours as monitored by ^{1}H-NMR (Scheme 1.17). Interfacing the library with lipase and isopropenyl acetate **104** as an acyl donor selected the cyanoacetate product **105** as the major product of the reaction.



Scheme 1.17 *Cyanohydrin DCLs targeting lipases*

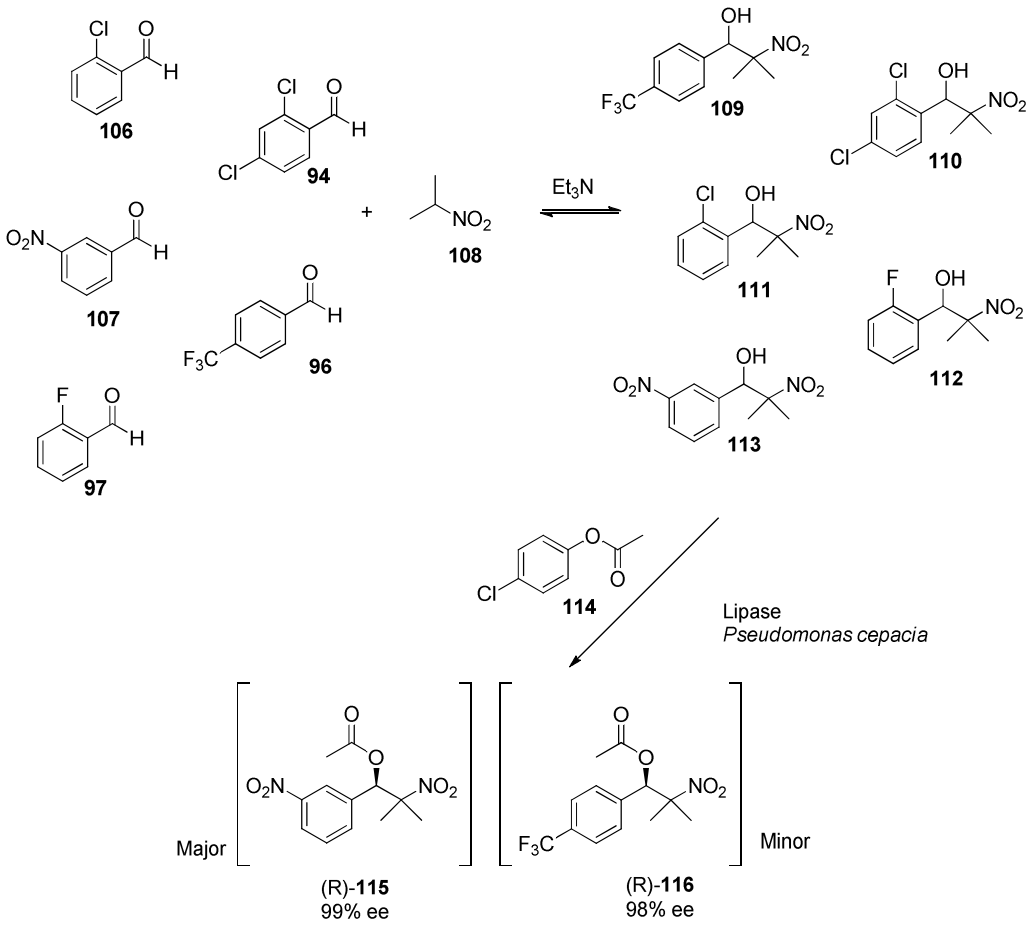
1.4.3 Nitroaldol DCR

The nitroaldol reaction, also known as the Henry reaction, was introduced to the field of DCR, again by Ramstrom and co-workers (Scheme 1.18).[49] It involves the nucleophilic addition of a nitroalkane to a carbonyl group forming a C-C bond. The reaction is thermodynamically controlled and can be accelerated in the presence of organic or inorganic bases or quaternary ammonium salts.



Scheme 1.18 *Nitroaldol DCR*

A ten-membered β-nitroalcohol DCl including all possible enantiomers was generated at 40^{o}C from equimolar amounts of five aromatic aldehydes and 2-nitropropane **108** (Scheme 1.19). 10 equivalents of triethylamine base were used to catalyse the library generation and equilibrium was established in 18 hours. Higher temperatures affected the enantioselectivity of the enzymatic reaction and also decomposed the library members.



Scheme 1.19 *Nitroaldol DCLs targeting lipases*

Selection of the lipase target required careful screening of a series of enzymes and acyl donors since lipase-catalysed transesterification of β-nitroalcohols had not been reported earlier in literature. The selected lipase *Pseudomonas cepacia* and five equivalents of the acyl donor, *p*-chlorophenyl acetate, were added to the nitroaldol library. Two of the β-nitroalcohols **109** and **113** were selectively transesterified to the corresponding acylated products. The overall yield for the two products was 24% in 24 hours and the reaction went to completion (95%) after 20 days. The DCR process was highly enantioselective with the R-enantiomer of the ester **115** resolved to 99% *ee* and the R-enantiomer of **116** to 98% *ee*. These examples established the conceptual framework of DCR as a valuable tool for the screening of a DCL of potential substrates using less than equimolar amounts of the target enzyme.

1.5 Target accelerated synthesis

Product distributions of reactions under kinetic control reflect the relative differences between the initial and transition state Gibbs energies and therefore biasing such kinetic processes would involve changing the energies of the transition states. This is generally achieved by preferential binding and stabilization of a selected transition state by an added template which would result in an increased rate of formation of the selected product. This method which leads to selection of the best binding molecule using irreversible reactions under kinetic control is referred to as target-accelerated synthesis (TAS) and is complementary to the thermodynamic approach (Figure 1.17).[50]

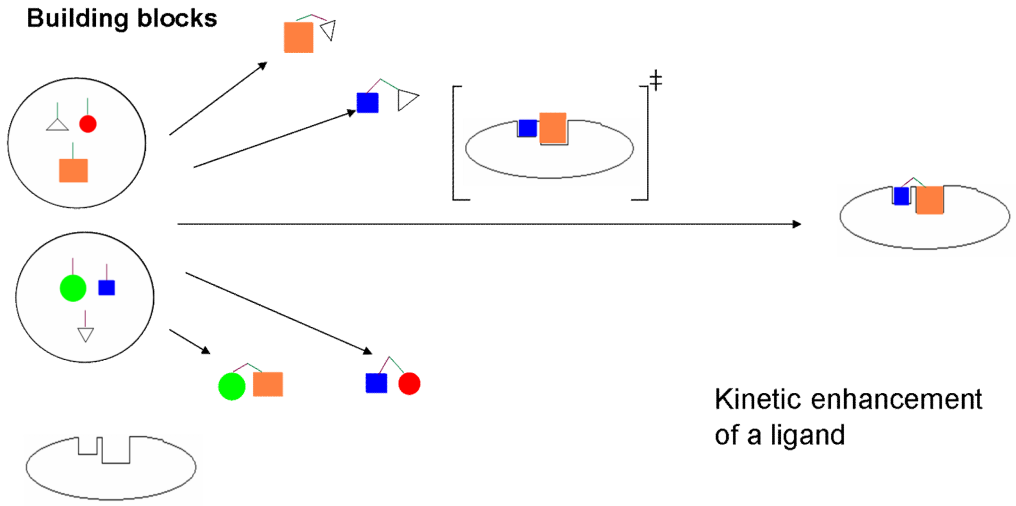
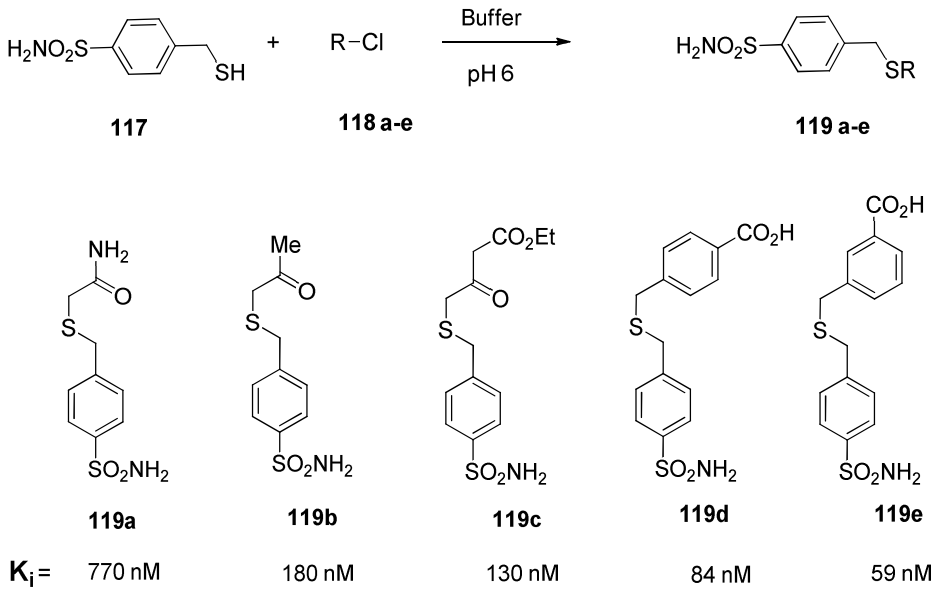


Figure 1.17 *Target-accelerated synthesis*

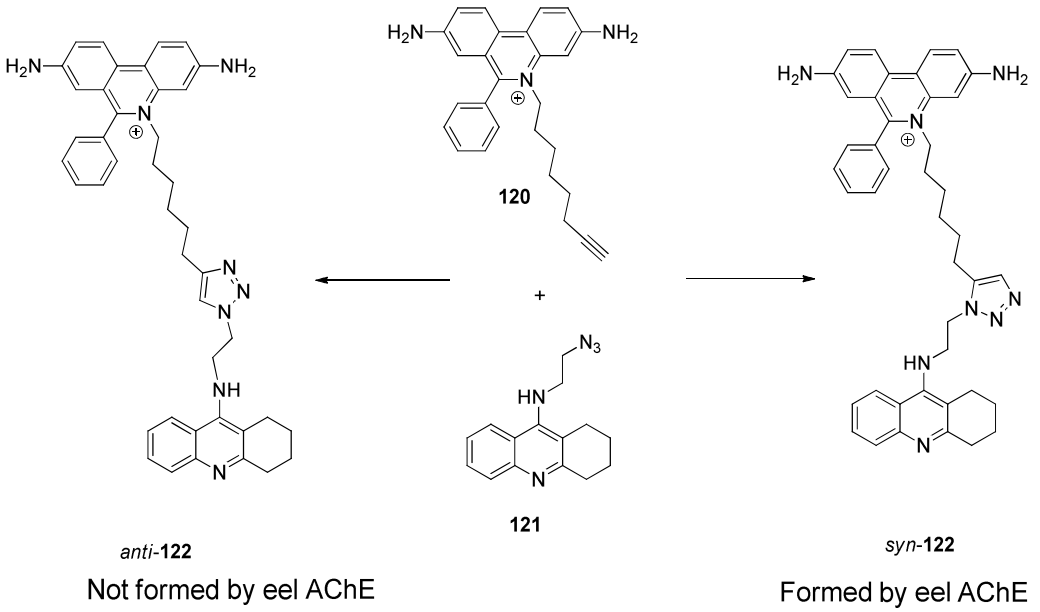
An elegant example was provided by Nguyen and Huc who used the active site of bovine carbonic anhydrase (BCA II) to template the formation of aromatic sulphonamide-based inhibitors through an irreversible nucleophilic substitution reaction (Scheme 1.20).[51] Products with a nine-fold difference in binding constants exhibited a two-fold increase in rate. Control experiments with bovine serum albumin (BSA), a protein that had no affinity for the building blocks, proved

conclusively that BCA II actually accelerates the formation of the best binding species.



Scheme 1.20 *Alkylation of thiols used to identify best binding ligands for BCA II*

The most successful target-accelerated synthesis method was developed by Sharpless who used the Huisgen 1,3-dipolar cycloaddition of alkynes and azides to form 1,4-disubsituted-1,2,3-triazoles as an ‘ in situ click’ reaction.[52-54] The reaction has many unique advantages for use in the presence of biological templates. The azide and alkyne functionalities are generally compatible with biomolecules and aqueous conditions while the triazole in the product is in many ways similar to the amide bond. The applicability of this reaction has been demonstrated for important target enzymes like carbonic anhydrase II and acetylcholinesterase (Scheme 1.21).[53,54]



Scheme 1.21 *In situ click chemistry targeting acetylcholinesterase*

For a successful target-accelerated synthesis experiment the target should simultaneously bind both the starting materials and also significantly accelerate the irreversible coupling reaction.[50] The approach would most likely fail in cases where two weakly binding species are linked together to form a strong binder. Optimising substituents around a tight- binding anchor scaffold would work only if the substituents themselves are relatively good binders. Such kinetic systems offer better selectivity than thermodynamic methods but are more difficult to control.

1.6 Conclusions and aims of this project

Protein-directed DCC is an emerging field which combines library synthesis and screening into a single step. The scope of the field depends heavily on the range of reversible synthetic chemistry that can be used under biocompatible conditions. C=N bond forming reactions were the first to be used to demonstrate the concept but lack true adaptability. Reversible disulfide formation, Michael addition of thiols to enones and enzymatic methods have attempted to provide alternatives to imine chemistry in

biological DCC. Pseudodynamic and DCR methods have been developed to overcome the requirement of stoichiometric amounts of protein targets, a constraint inherent to the DCC concept. As this field evolves new ways of generating biocompatible dynamic libraries are needed in order to expand the structural diversity accessible to DCC.

This aim of this project is to develop new methods for adaptive DCLs which can be readily interfaced with protein targets. Our experiments establish catalysed acyl hydrazone formation as a new strategy for generating DCLs targeting proteins. Chapter II of this thesis deals with the development of this methodology and its application for the identification of isoform-specific binders for two different GST enzymes. Chapter III outlines our efforts towards developing multi-level DCLs which use two orthogonal reversible reactions as an entry point into the field of systems chemistry.

1.7 Notes and References

1 Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J-L.; Sanders, J. K. M.; Otto, S. *Chem. Rev.* **2006**,*106*, 3652-3711.

2 Dynamic Combinatorial Chemistry in Drug Discovery,Bioorganic Chemistry, and Materials Science; edited by Miller, B.L.; Wiley, 2010.

3 Huc, I.; Lehn, J.-M. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 2106-2110.

4 Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Curr. Opin. Chem. Biol.* **2002**, *6*, 321-327.

5 Fuji, S.; Lehn, J.-M. *Angew. Chem., Int. Ed.* **2009**, *48*, 7635-7638.

6 Hagihara, S.; Tanaka, H.; Matile, S. *J. Am. Chem. Soc.* **2008**, *130*, 5656-5658.

7 Corbett, P. T.; Sanders, J. K. M.; Otto, S. *Angew. Chem., Int. Ed.* **2007**, *46*, 8858- 8861.

8 Ramström, O.; Lehn, J.-M. *Nat. Rev. Drug Discovery* **2002**, *1*, 26-36.

9 Lehn, J.-M. *Chem. Eur. J.* **1999**, *3*, 2455-2463.

10 Bunyapaiboonsri, T.; Ramström, O.; Lohmann, S.; Lehn, J.-M.; Peng, L.; Goeldner, M. *ChemBioChem* **2001**, *2*, 438-444.

11 Cousins, G. R. L.; Poulsen, S.-A.; Sanders, J. K. M. *Chem. Commun.* **1999**, 1575-1576.

12 Bunyapaiboonsri, T.; Ramström, H.; Ramström, O.; Haiech, J.; Lehn, J.-M. *J. Med. Chem.* **2003**, *46*, 5803- 5811.

13 Hochgürtel, M.; Kroth, H.; Piecha, D.; Hofmann, M. W.; Nicolau, C.; Krause, S.; Schaaf, O.; Sonnenmoser, G.; Eliseev, A. V. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *9*, 3382-3387.

14 Hochgürtel, M.; Biesinger, R.; Kroth, H.; Piecha, D.; Hofmann, M. W.; Krause, S.; Schaaf, O.; Nicolau, C.; Eliseev, A. V. *J. Med. Chem.* **2003**, *46*, 356-358.

15 Zameo, S.; Vauzeilles, B.; Beau, J.-M. *Angew. Chem., Int. Ed.* **2005**, *44*, 965-969.

16 Zameo, S.; Vauzeilles, B.; Beau, J.-M. *Eur. J. Org. Chem.* **2006**, 5441-5444.

17 Valade, A.; Urban, D.; Beau, J.-M. *ChemBioChem* **2006**, *7*, 1023-1027.

18 Valade, A.; Urban, D.; Beau, J.-M. *J. Comb. Chem.* **2007**, *9*, 1-4.

19 Congreve, M. S.; Davis, D. J.; Devine, L.; Granata, C.; O'Reilly, M.; Wyatt, P. G.; Jhoti, H. *Angew. Chem.,Int. Ed.* **2003**, *42*, 4479-4482.

20 For an early example of using X-ray crystallography to identify binding events templated by a protein, see: Katz, B. A.; Finer-Moore, J.; Mortezaei, R.; Rich, D. H.; Stroud, R. M. *Biochemistry* **1995**, *34*, 8264-8280.

21 Shahed, Z.; Whitesides, G.M. *Biochemistry*, **1980**,*19*,4156-4166.

22 Hioki, H.; Still, W. C. *J. Org. Chem.* **1998**, *63*, 904-905.

23 Ramström, O.; Lehn, J.-M. *ChemBioChem* **2000**, *1*, 41-48.

24 Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *J. Am. Chem. Soc.* **2000**, *122*, 12063-12064.

25 Ramström, O.; Lohmann, S.; Bunyapaiboonsri, T.; Lehn, J.-M. *Chem.--Eur. J.* **2004**, *10*, 1711-1715.

26 Milanesi, L.; Hunter, C. A.; Sedelnikova, S. E.; Waltho, J. P. *Chem. Eur. J.* **2006**, *12*, 1081-1087.

27 Danieli, B.; Giardini, A.; Lesma, G.; Passarella, D.; Peretto, B.; Sacchetti, A.; Silvani, A.; Pratesi, G.; Zunino,F. *J. Org. Chem.* **2006**, *71*, 2848-2853.

28 Scott, D.E.; Dawes, G.J.; Abell,C.;Ciulli, A. *ChemBioChem* **2009**,*10*, 2772 – 2779

29 Erlanson, D. A.; Ballinger, M. D.; Wells, J. A. *Fragment-based approaches in drug discovery*, Eds. Jahnke,W.; Erlanson, D. A., Wiley-VCH, Weinheim, 285-308.

30 Braisted, A. C.; Oslob, J. D.; Delano, W. L.; Hyde, J.; McDowell, R. S.; Waal, N.; Yu, C.; Arkin, M. R.;Raimundo B. C. *J. Am. Chem. Soc.* **2003**, *125*, 3714–3715.

31 Erlanson, D. A.; Lam, J. W.; Wiesmann, C.; Luong, T. N.; Simmons, R. L.; Delano, W. L.; Choong, I. C.;Burdett, M. T.; Flanagan, W. M.; Lee, D.; Gordon, E. M.; O'Brien, T. O. *Nat. Biotechnol.* **2003**, *21*, 308-314.

32 Hardy, J. A.; Lam, J.; Nguyen, J. T.; O'Brien, T.; Wells, J. A. *Proc. Nat. Acad. Sci. USA*, **2004**, *101*, 12461-12466.

33 Buck. E.; Bourne, H.; Wells, J. A. *J. Biol. Chem.* **2005**, *280*, 4009-4012.

34 Buck. E.; Wells. J. A. *Proc. Nat. Acad. Sci USA* **2005**, *102*, 2719-2724.

35 Erlanson, D. A.; Braisted, A. C.; Raphael, D. R.; Randal, M.; Stroud, R. M.; Gordon, E. M.; Wells, J. A. *Proc.Nat. Acad. Sci. USA* **2000**, *97*, 9367-9372.

36 Caraballo, R.; Dong, H.; Rebeiro, J.P.; Jinemez-Barbero, J.;Ramström, O. *Angew. Chem.,Int. Ed.* **2010**, *49*, 589-593.

37 Swann, P. G.; Casanova, R. A.; Desai, A.; Frauenhoff, M. M.; Urbancic, M.; Slomczynska, U.; Hopfinger, A.J.; Le Breton, G. C.; Venton, D. L. *Biopolymers*, **1997**, *40*, 617-625.

38 Hoerter, J. M.; Otte, K. M.; Gellman, S. H.; Cui, Q.; Stahl, S. S. *J. Am. Chem. Soc.* **2008**, *130*, 647-654.

39 Eldred, S. E.; Stone, D. A.; Gellman, S. H.; Stahl, S. S. *J. Am. Chem. Soc.* **2003**, *125*, 3422-3423.

40 Bell, C. M.; Kissounko, D. A.; Gellman, S. H.; Stahl, S. S. *Angew. Chem., Int. Ed.* **2007**, *46*, 761-763.

41 Lins, R. J.; Flitsch, S. L.; Turner, N. J.; Irving, E.; Brown, S. A. *Angew. Chem., Int. Ed.* **2002**, *41*, 3405-3407.

42 Lins, R. J.; Flitsch, S. L.; Turner, N. J.; Irving, E.; Brown, S. A. *Tetrahedron* **2004**, *60*, 771-780.

43 Cheeseman, J. D.; Corbett, A. D.; Shu, R.; Croteau, J.; Gleason, J. L.; Kazlauskas, R. J. *J. Am. Chem. Soc.***2002**, *124*, 5692-5701.

44 Corbett, A. D.; Cheeseman, J. D.; Kazlauskas, R. J.; Gleason, J. L. *Angew. Chem.,Int. Ed.* **2004**, *43*, 2432-2436.

45 Shi, B.; Greaney, M. F. *Chem. Commun.* **2005**, 886-888.

46 Shi, B.; Stevenson, R.; Campopiano, D. J.; Greaney, M. F. *J. Am. Chem. Soc.* **2006**, *128*, 8459-8467.

47 Larsson, R.; Pei, Z.; Ramström, O. *Angew. Chem., Int. Ed.* **2004**, *43*, 3716-3718.

48 Larsson, R.; Ramström, O. *Eur.J. Org.Chem.* **2005**,285-291

49 Vongvilai, P.; Angelin, M.; Larsson, R.; Ramström, O. *Angew. Chem. Int. Ed.* **2007**, *46*, 948-950.

50 Cheeseman, J.D.; Corbett, A. D.; Cleason, J.L.; Kazlauskas, R.J. *Chem. Eur. J.* **2006**, *11*, 1708-1716.

51 Nguyen, R.; Huc, I. *Angew. Chem., Int. Ed.* **2001**, *40*, 1774-1776.

52 Whiting, M.; Muldoon. J.; Lin, Y.C.; Silverman,S.M.; Lindstrom, W.; Olson, A.J.; Kolb, H.C.; Finn, M.G.; Sharpless, K.B.; Elder,J.H.; Fokin, V.V. *Angew. Chem., Int. Ed.* **2006**, *45*, 1435-1439.

53 Mocharla, V.P.; Colasson, B.; Lee, L.V.; Röper, S.; Sharpless K.B.; Wong C.H.; Kolb H.C. *Angew.Chem., Int. Ed.* **2004**, *44*, 116-120.

54 Lewis,W.G.; Green, L.G.; Grynszpan, F.; Radić, Z.; Carlier, P.R.; Taylor, P.; Finn, M.G.; Sharpless, K.B. *Angew.Chem., Int. Ed.* **2002**, *41*,1053-1057.

Chapter 2

Catalysing Dynamic Acylhydrazone Libraries

2.1 Introduction

A critical review of literature indicates that the field of dynamic combinatorial chemistry has the potential to develop into an extremely useful tool for exploring protein-ligand interactions even in the absence of extensive structural information on the protein binding site. The main challenges in implementing DCC with biological templates lie in performing the synthetic chemistry under conditions which maintain the structural and functional integrity of the biomolecule. The reversible reactions must take place under mild physiological conditions, eliminating those that are incompatible with water. The process should not interfere with either functional groups present on the bio-template or the molecular recognition involved in the selection process. The exchange process should attain equilibrium within a reasonable timeframe in which the structural and functional integrity of the template is preserved. Since very few reversible reactions meet these strict requirements of biocompatibility, Lehn proposed two limiting protocols for constructing DCLs targeting biomolecules: adaptive and pre-equilibrated.[1] The adaptive DCL involves a bio-compatible reversible reaction so that library generation and selection of the best binders can be performed in a single step in the presence of the target. The DCL may be real or virtual but the binding events involving the target control the evolution of the library composition. The adaptive approach usually requires stoichiometric amounts of the target for appreciable amplification to be observed.

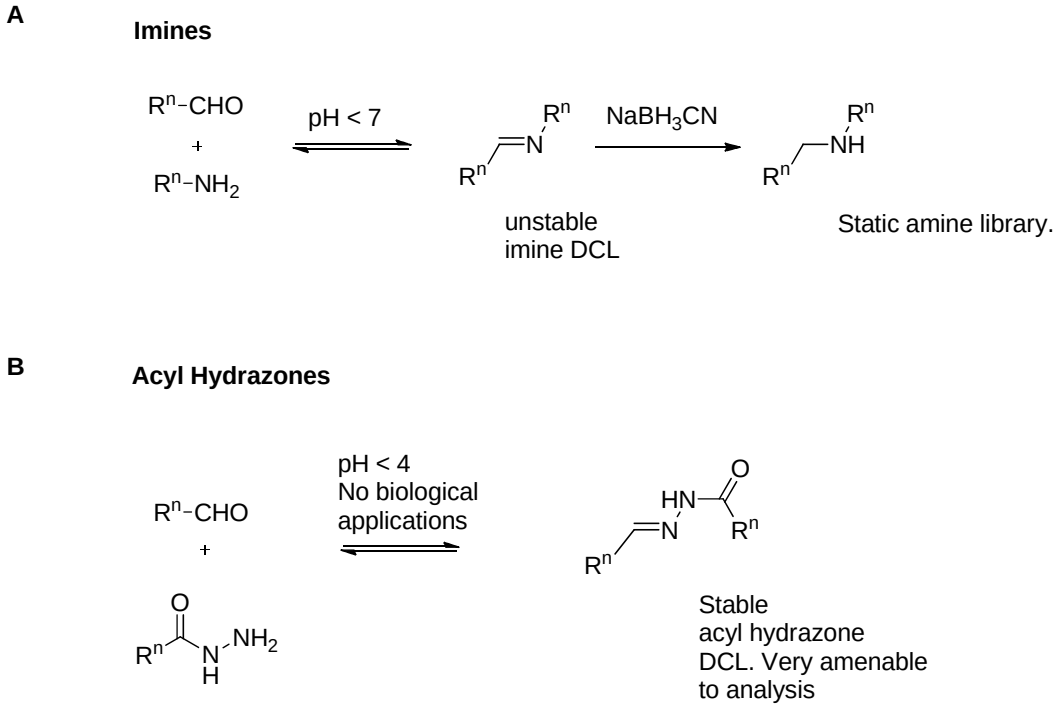
In the case of pre-equilibrated DCLs the reversible reaction is not compatible with the biomolecular template. Therefore the dynamic library is generated in the absence of the target which is introduced into the system only after the exchange process has been switched off. This compromises the adaptive feature of DCC and is akin to the post-synthesis screening of a static combinatorial library. An advantage of this method is that the screening process in the second step does not require stoichiometric amounts of the target. This is important only when small quantities of target protein are available or if it is very expensive.

Methods for generating truly adaptive DCLs targeting proteins are limited and are mostly based on the disulfide exchange[2] or thiol conjugate addition[3,4] which have been reviewed in Chapter I of this thesis. This lack of reaction diversity is a major

constraint, limiting the use of this powerful concept for drug discovery applications. The success of the DCC approach to studying biological systems, therefore, depends on the development of new tools for generating adaptive biocompatible DCLs. This chapter deals with our attempts towards the development of new methods for adaptive DCLs which can be readily interfaced with protein targets. We have identified nucleophilic catalysis of acylhydrazone equilibration using aniline as a promising approach to generate truly adaptive DCLs under near-physiological conditions.

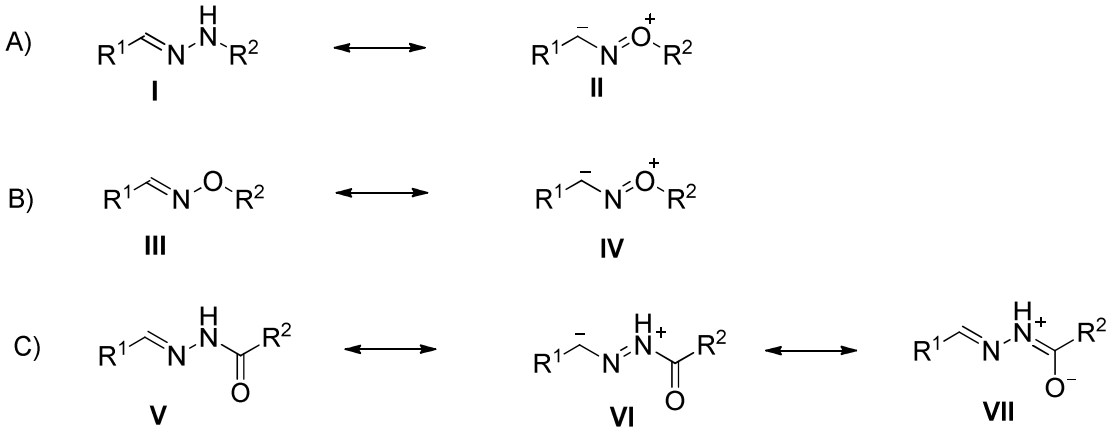
2.2 The reaction of choice: Acylhydrazone exchange

Reversible amine/carbonyl condensations generating imines, hydrazones and oximes are ideal candidates for biological DCC applications given the prevalence of imine chemistry in biochemical systems.[5] Simple imines form rapidly but present serious analytical and isolation problems due to their inherent instability in aqueous solutions (Scheme 2.1A). With hydroxylamines and acyl hydrazides the equilibrium lies towards the products with relatively stable oximes and hydrazones being formed but the kinetics is much slower especially under physiological conditions.[6] The reactions ideally require acid catalysis (pH < 4) and are frozen under neutral or basic conditions (Scheme 2.1B).[7,8]



Scheme 2.1 *Transimination reactions for DCC.* ***A:*** *Imine DCLs- reversible addition of amines to aldehydes gives unstable imines that cannot be isolated or analysed directly, necessitating an in situ reduction step. The resultant static library of amines may or may not share the binding profile of the imine precursors.* ***B:*** *Acyl hydrazone DCLs- reaction of aldehydes with hydrazides gives acylhydrazones that have good stability and are amenable to analysis. Equilibration requires acidic conditions that are incompatible with a biological target.*

The greater intrinsic hydrolytic stability of hydrazones and oximes over imines is believed to be due to electron delocalisation (Scheme 2.2). Resonance forms IV and VI contribute towards increasing the negative charge density on the imine carbon thereby reducing its electrophilicity.[9]



Scheme 2.2 *Hydrolytic stability of hydrazones and oximes*

Acylhydrazone exchange was introduced to DCC by the Sanders group who applied it to a large number of abiotic DCC systems.[10-12] However this reaction has not found much use for adaptive biological DCC due to the acidic conditions required for equilibration in a reasonable timeframe. An exception is the work by Poulsen who has shown that slow equilibration of acylhydrazones, taking one week at pH 7.2, can enable the identification of binders for the enzyme carbonic anhydrase using mass spectrometry.[13] The acylhydrazone functionality has unique features; the amide bond provides hydrogen bonding sites important for protein templates while the imino bond is responsible for the dynamic character through reversibility (Figure 2.1).[14] We envisaged the use of this reaction for generating adaptive DCC systems with protein templates if a suitable catalyst could be found to accelerate the exchange process under physiological conditions.

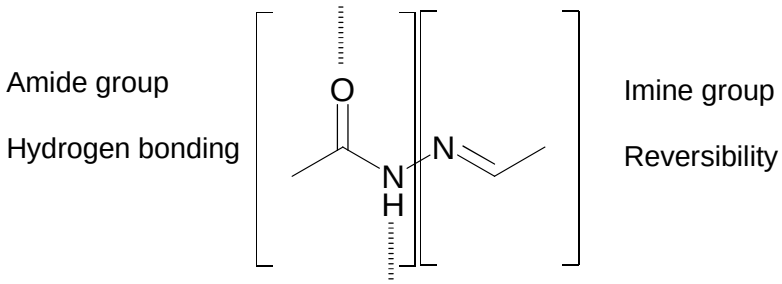
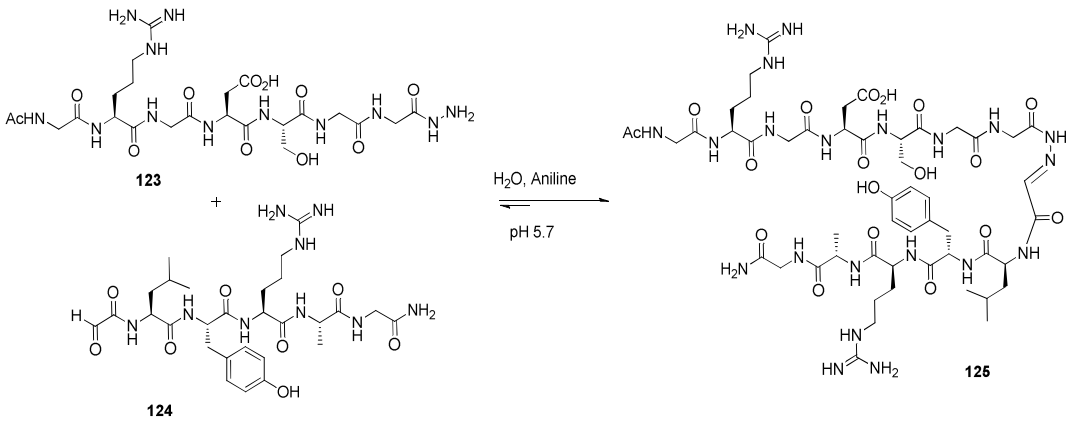


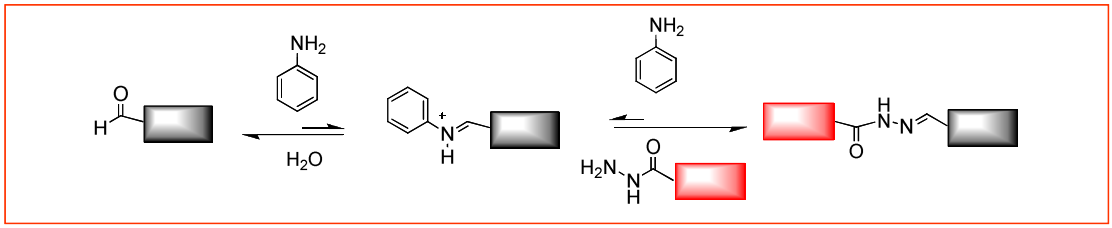
Figure 2.1 *The Acylhydrazone linkage*

Classic work from Jencks in the 1960s led to the identification of aniline derivatives as nucleophilic catalysts for semicarbazone formation[15] which was recently extended to hydrazone and oxime formation in peptide ligation systems by Dawson (Scheme 2.3A).[16-18] Aniline-catalysed transimination proceeds through aniline Schiff base intermediates which are highly unstable in aqueous conditions (Scheme 2.3B).[16] This allows for large concentrations of the catalyst to be used, facilitating fast equilibration of hydrazones and oximes even in millimolar concentrations. Consistent with nucleophilic catalysis, aniline does not affect the overall equilibrium constant of the reaction and thus maintains equilibrium distribution of reactants and products. In this project we explore the use of aniline to generate acylhydrazone DCLs under conditions favourable for proteins.

A



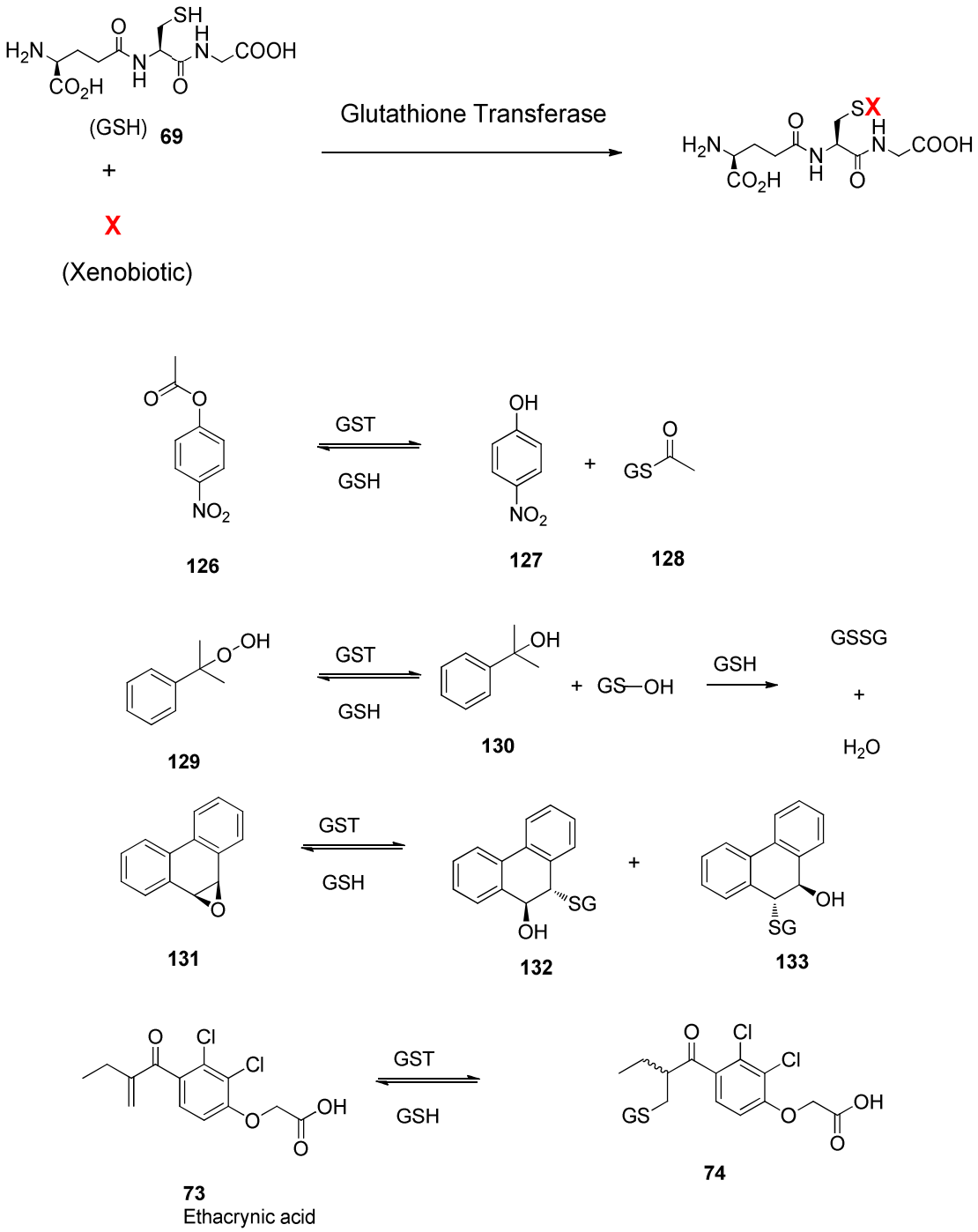
B



Scheme 2.3 *Aniline catalysis of acyl hydrazone formation and exchange. A: Aniline catalysis used by Dawson for peptide ligation systems. B: A highly electrophilic transient aniline Schiff base accelerates acylhydrazone exchange*

2.3 The target protein: Glutathione S-tranferases

The Glutathione Transferases (EC 2.5.1.18, GSTs) are detoxification enzymes which catalyse the conjugation of glutathione (GSH) **69**, to a wide variety of hydrophobic electrophilic substrates (Scheme 2.4).[19,20]



Scheme 2.4 *Reactions catalysed by the GSTs*

Based on their substrate specificity and primary structure cytosloic GSTs are divided into five different classes (alpha, mu, pi, sigma and theta).[21,22] The enzymes exist as homo- or heterodimers with each monomer having its own active site consisting of a GSH-binding site (G-site) and a hydrophobic binding site (H-site) (Figure 2.2).[23] The G-site is conserved across different isoforms but the H-site is highly diverse and responsible for the large variety of substrates recognized by the GSTs. The catalytic role of the GSTs is dependent on highly conserved tyrosine, serine or cysteine residues at the G-site.[24] Upon binding to the enzyme, the pK_a of GSH drops from ~ 9.5 to ~ 6.8 units which aids deprotonation to form the more nucleophilic thiolate anion under physiological conditions. In addition to these two sites some GSTs have a third ligandin-binding site (L-site) which is able to bind large hydrophobic molecules like steroids, bile salts, fatty acids or drugs.[25] The exact location of this L-site is uncertain.

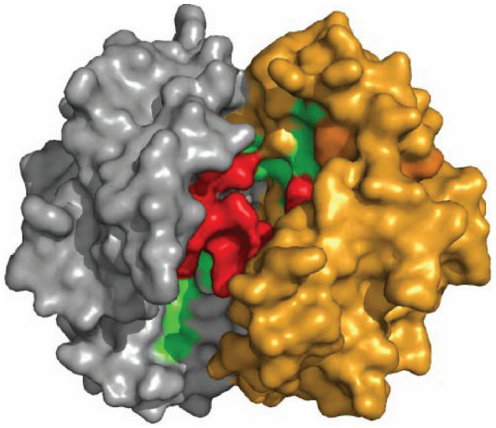
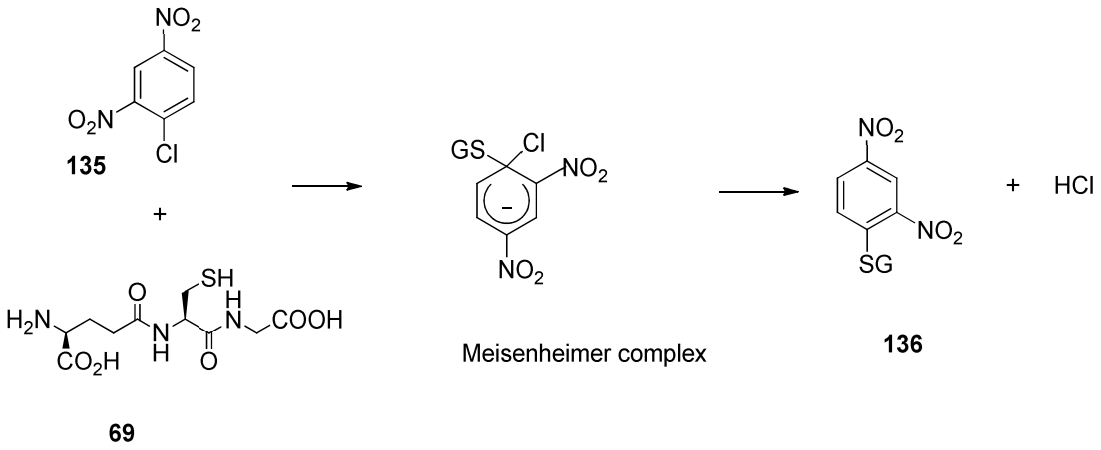


Figure 2.2 *Structure of GST that illustrates the H- and the G- sites. (Grey: monomer 1; yellow: monomer 2; green: G site; red: H site)*

These well-characterised proteins are relatively stable and easy to purify. They can be easily assayed using the 'CDNB assay' which uses the aromatic nucleophilic substitution reaction between GSH and 1-chloro-2,4-dinitrobenzene (CDNB, **135**) considered to be a universal GST substrate (Scheme 2.5).[19]



Scheme 2.5 *The CDNB assay for GSTs*

The GSTs are considered to be important drug targets in cancer since resistance to chemotherapeutic drugs has been directly correlated with overexpression of GSTs in tumour cells.[26] Previous work in the Greaney group involved the development of thiol conjugate addition DCLs for identifying GST inhibitors.[4] The bisubstrate architecture of these proteins makes them targets eminently suitable for a fragment-based DCC approach since the concept relies on reversibly connecting two sets of fragments together to generate a dynamic library.[1] In this project, we have focussed on two recombinant GST isozymes as targets, SjGST from the helminth worm *Schistosoma japonicum,* a drug target in tropical disease,[27] and hGST P1-1, a human isoform important in cancer.[26]

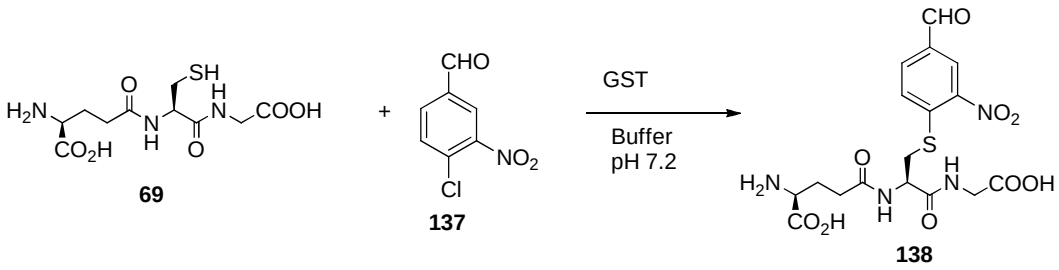
2.4 Results and discussion

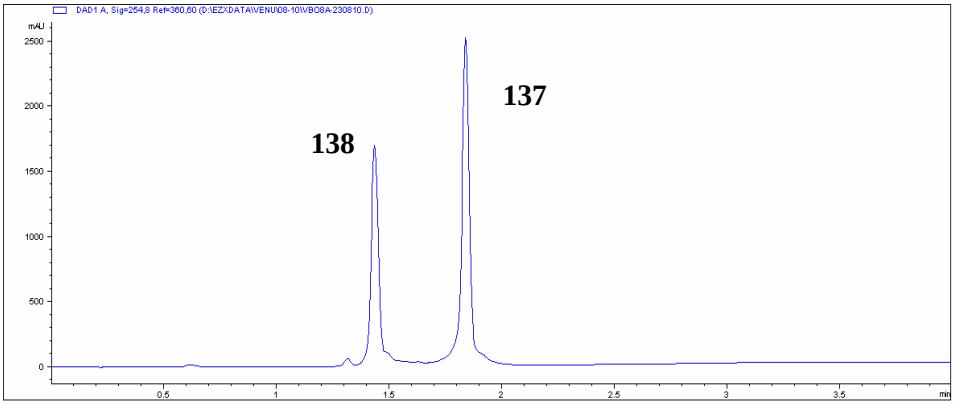
2.4.1 Design of the dynamic acylhydrazone library

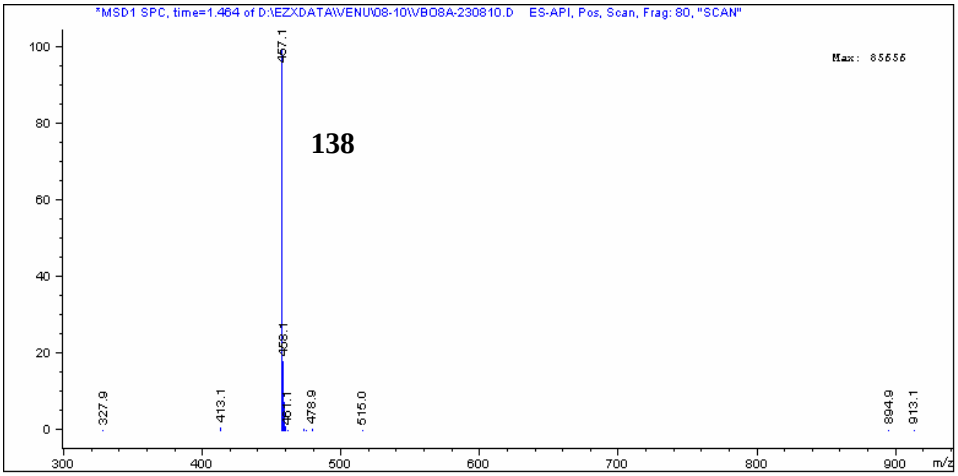
A critical parameter in the design of a DCL targeting proteins is the selection of the building blocks which undergo reversible exchange. At least some of the building blocks should have characteristic molecular features that can be recognised by and interact with the target binding site. We envisaged an n x 1 DCL system which uses a biologically active scaffold that can be derivatized at a single position with n components. In comparison to n x n DCLs the combinatorial aspect is diminished in

the n x 1 case generating only n + 1 components in the library. However amplification in this case can be very strong and, in principle, the entire library can be amplified to a single member. The combinatorial size of the library can be improved by using multivalent scaffolds in the n + 1 DCLs without sacrificing the amplification.

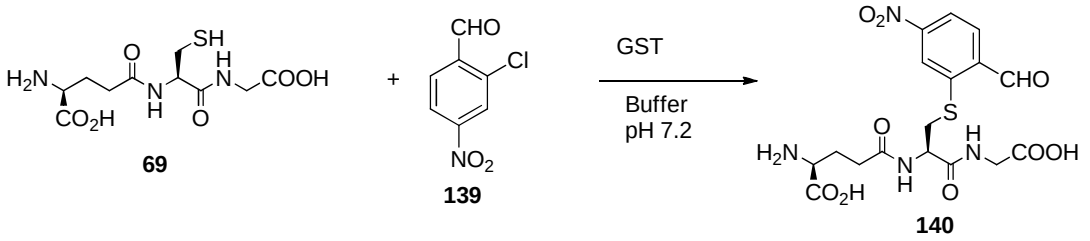
Our design of the DCC experiments involved using a scaffold aldehyde **137** structurally related to the known GST substrate, chlorodinitrobenzene (CDNB) **135**. We expected this structure to have some affinity for the enzyme H-site. Upon incubating this aldehyde **137** with 1 equivalent of GSH and catalytic amounts of SjGST in ammonium acetate buffer (50 mM, pH 7.2) we identified the glutathione conjugate aldehyde **138** formed by LC-MS (Scheme 2.6). Interestingly, aldehyde **139**, a regioisomer of **137**, also gave glutathione conjugates with the enzyme (Scheme 2.7). Control experiments under similar conditions in the absence of GST did not produce any glutathione conjugates for both the aldehydes indicating that they act as substrates for the protein and access the H-sites. Experiments repeated with the human isoform hGSTP1-1, gave very similar results. (Both enzyme samples were received from Anne Caniard and Dominic Campopiano)

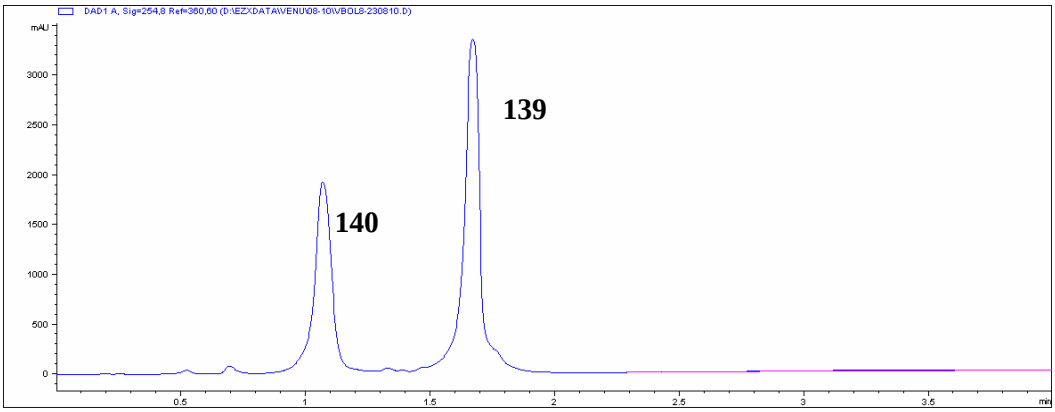


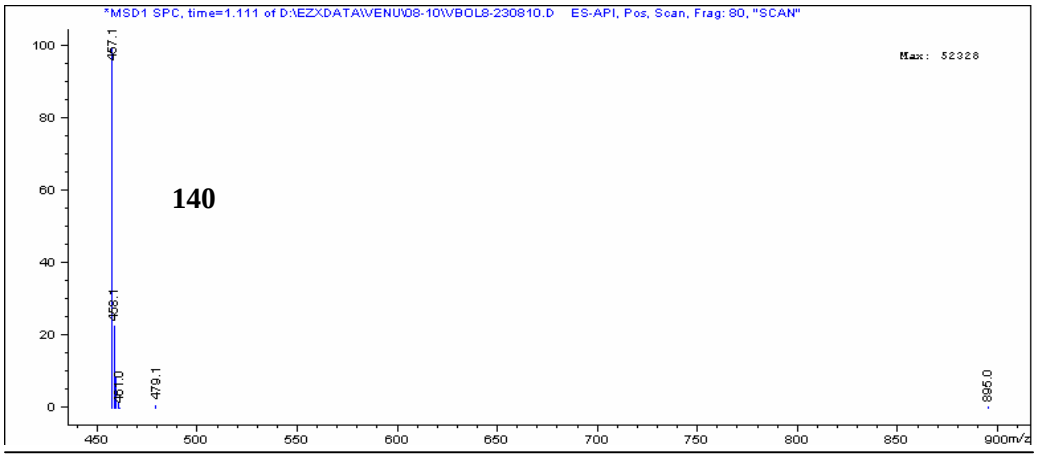




Scheme 2.6 *The GSTs catalyse glutathione conjugation of aldehyde* ***137****(LC-MS data)*



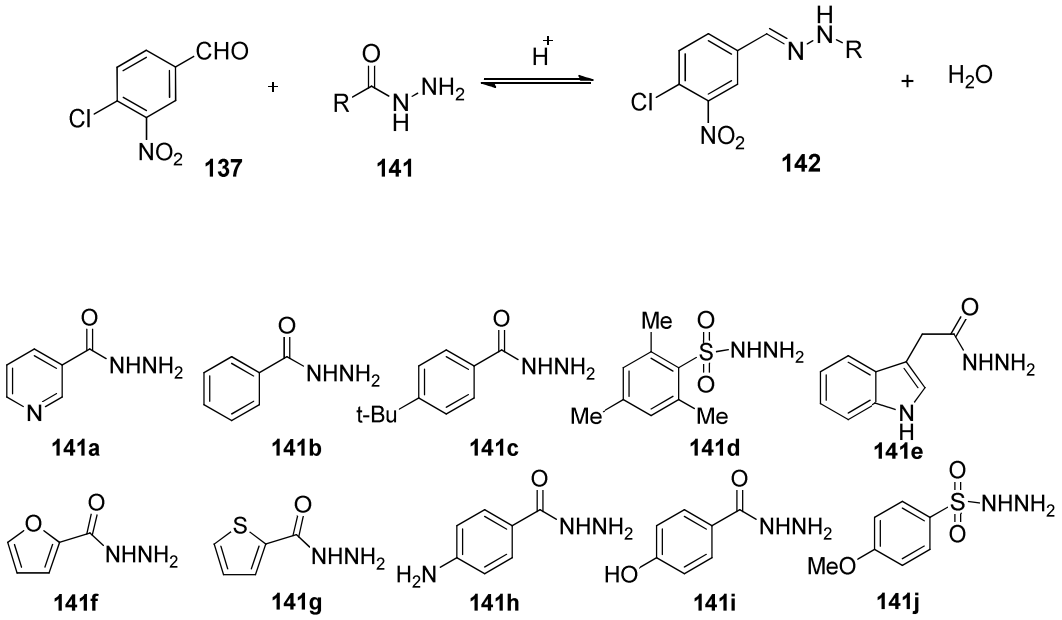




Scheme 2.7 *The GSTs catalyse glutathione conjugation of aldehyde* ***139****(LC-MS data)*

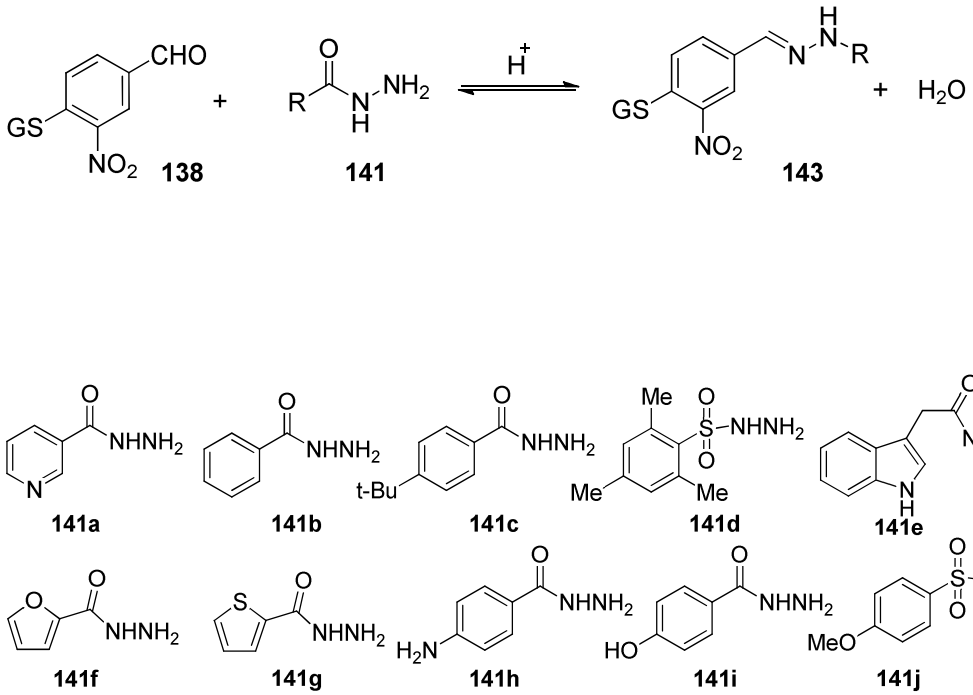
To generate the DCL with the aldehyde **137** we randomly chose ten commercially available aryl hydrazides (Scheme 2.8). This library consisted of eight acyl and two sulfonyl hydrazides displaying a range of aryl and heteroaryl groups. We reasoned

that this diverse set of acylhydrazones bearing a common chloronitrobenzaldehyde moiety would be able to access the hydrophobic pocket at the GST binding site under suitable conditions. The hydrazones would also have similar UV absorption properties to allow sensitive detection at a common wavelength.

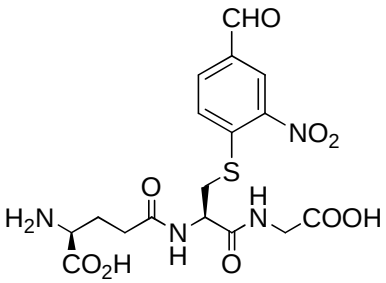


Scheme 2.8 *Acyl hydrazone DCL based on ten selected hydrazides*

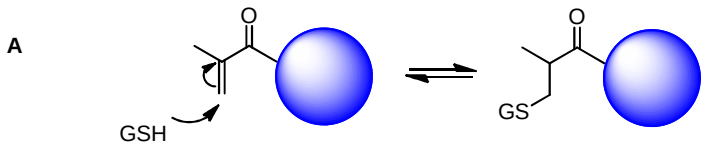
To increase the potency of our DCL components we envisaged the use of aldehyde **138** as our core scaffold to generate an n x 1 DCL with the same hydrazides **141a-j** (Scheme 2.9). We expected the glutathione motif of the aldehyde to act as an 'anchor' at the enzyme G-site thereby enabling the exploration of the H-site with the different hydrazide fragments.[28] The γ-glutamyl residue of GSH is known to be critical for its binding at the hydrophilic G-site forming a number of key hydrogen bonds and salt bridges.[29-31] This 'anchoring' approach uses GSH in a passive role in strict contrast to previous work in the Greaney group based on conjugate addition of thiols to enones,[4] where GSH played an active role in the reversible reaction used to generate the DCL (Scheme 2.10).

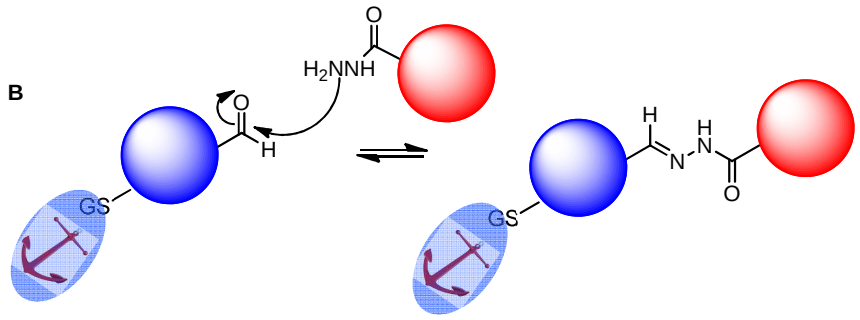


Scheme 2.9 *Glutathione-conjugated acylhydrazone DCL based on ten selected hydrazides*



138

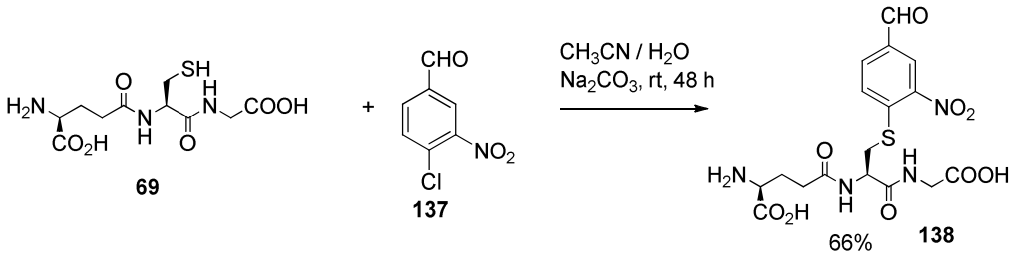




Scheme 2.10 *The concept of anchoring. A: reversible Michael addition of thiols to enones where GSH plays an active role as a reactant. B: the GS- moiety is 'anchored, on to the enzyme G-site.*

2.4.2 Synthesis of glutathione conjugated aldehyde

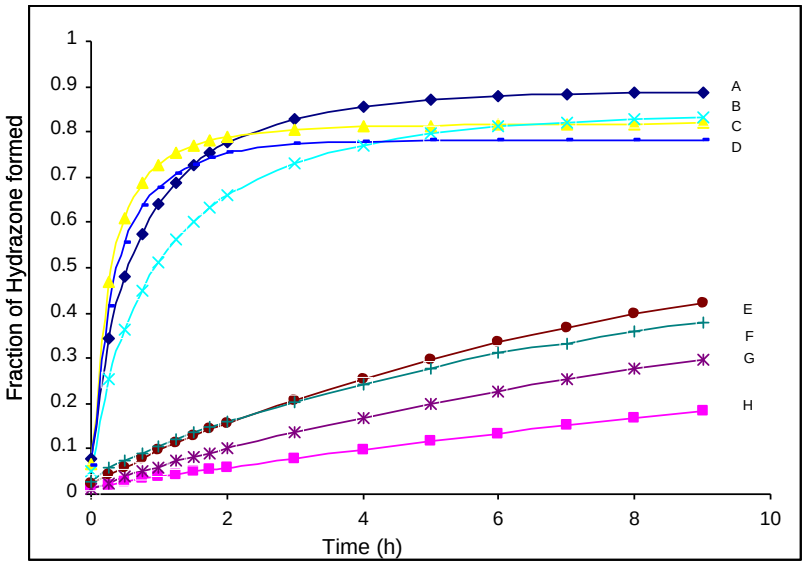
Glutathione conjugate aldehyde **138** was synthesised by an S_NAr reaction under basic conditions in a 1:1 mixture of MeCN and water as outlined in Scheme 2.11. After 48 hours stirring at room temperature, the yellow solution was washed with diethyl ether, concentrated and then purified by preparative HPLC.



Scheme 2.11 *Synthesis of glutathione conjugates aldehyde scaffold*

2.4.3 Aniline catalysis of acylhydrazone formation

To optimize aniline catalysis for our DCLs we first decided to probe if any substituents on aniline improve the rate of hydrazone formation and exchange. We measured the initial rates of reaction between aldehyde **138** (0.5 mM, aqueous) and hydrazide **141b** (0.5 mM, DMSO) for a range of aniline derivatives **A, B, C, D E, G** and imidazole **F** (Figure 2.3). The reactions performed at ambient temperature in ammonium acetate buffer (0.1M, pH 6.2, 10% DMSO) were monitored by UV-HPLC at 254 nm. Interestingly anilines (at concentrations of 10 mM) with electron-donating groups **(A,B,C,D)** increase the rate with $t_{1/2}$ values < 30 minutes while electron-poor and sterically hindered anilines show $t_{1/2}$ values >> 10 hours (graph in Figure). Since rate enhancement by aniline **(A)** was comparable to *p*-methoxyaniline **(B)** and aniline is more compatible with aqueous environments we decided to use aniline for catalysing our DCLs.



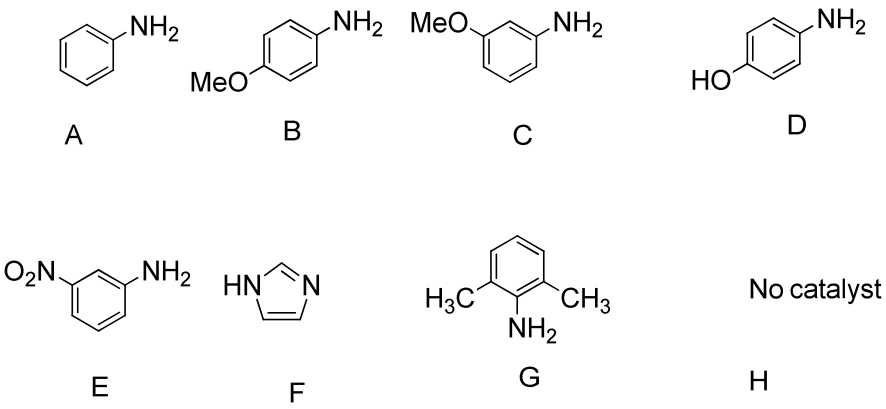


Figure 2.3 *Substituted anilines and imidazole catalysing the acyl hydrazone reaction: the graph shows the formation of hydrazone* ***143b*** *over time at a reactant concentration of 0.5mM each in 0.1M ammonium acetate buffer (pH 6.2) in the absence and in the presence of different substituted anilines (10 mM). Peak integrals from RP-HPLC traces (254 nm) used to measure the fraction of hydrazones formed*

2.4.4 Effect of aniline concentration on DCL equilibration: Single hydrazone formation

We further optimised the amount of the nucleophilic catalyst required by measuring the rate of formation of hydrazone **143b** at different aniline concentrations ranging from 0.01-10 mM (Figure 2.4). At individual reactant concentrations of 1 mM in buffer at pH 6.2, significant acceleration of the reaction rate was observed for aniline concentrations in the range 1-10 mM. A low concentration of 0.01 mM aniline did not significantly alter the rate in comparison to the uncatalysed reaction. These

relatively large concentrations of aniline achieve high rates of hydrazone formation due to the *in situ* generation of high concentrations of the electrophilic aniline Schiff base.[16]

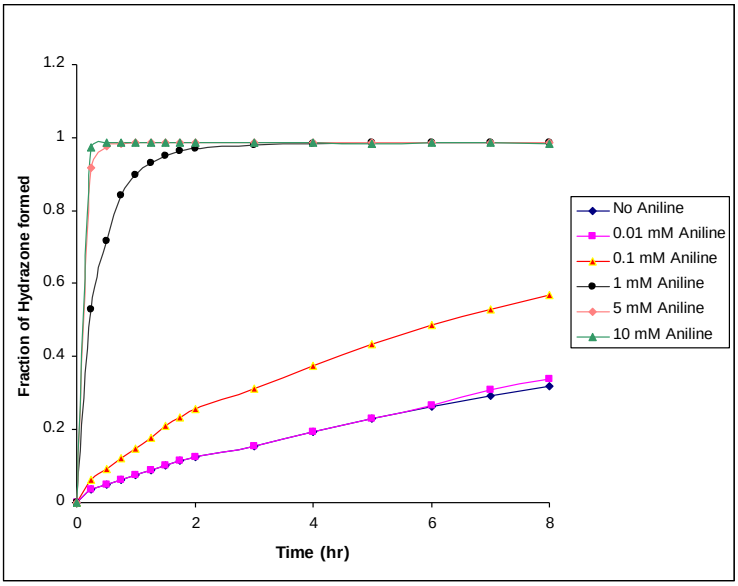


Figure 2.4 *Effect of aniline concentration on the catalysis of acylhydrazone reaction: the graph shows the formation of hydrazone* ***143b*** *over time at a reactant concentration of 1mM each in 0.1M ammonium acetate buffer (pH 6.2) in the absence and in the presence of different concentrations of aniline. Peak integrals from RP-HPLC traces (254 nm) were used to measure the fraction of hydrazones formed*

2.4.5 Effect of aniline concentration on DCL equilibration: 3 membered DCL

To adapt the results of our studies on aniline catalysis of acylhydrazone formation for generating biocompatible DCLs we started with small 3-membered DCLs formed from aldehyde **138** and an equivalent each of the hydrazides **141b, 141e** and **141f**. We set up the libraries at four different aniline concentrations (10 mM, 1 mM, 0.1 mM, and 0.01 mM) in buffer/DMSO mixtures (pH 6.2) and analysed their composition by HPLC over different time intervals. The libraries with 10 mM and 1 mM aniline equilibrated very quickly with a steady state composition achieved in less than 4 hours (Figures 2.5a and b).

On the other hand the DCLs with 0.1 mM and 0.01 mM aniline took 24 – 48 hours for complete equilibration (Figures 2.5c and d). We thus concluded that an optimum

catalyst concentration in the millimolar range would be required for generating our larger hydrazone DCLs.

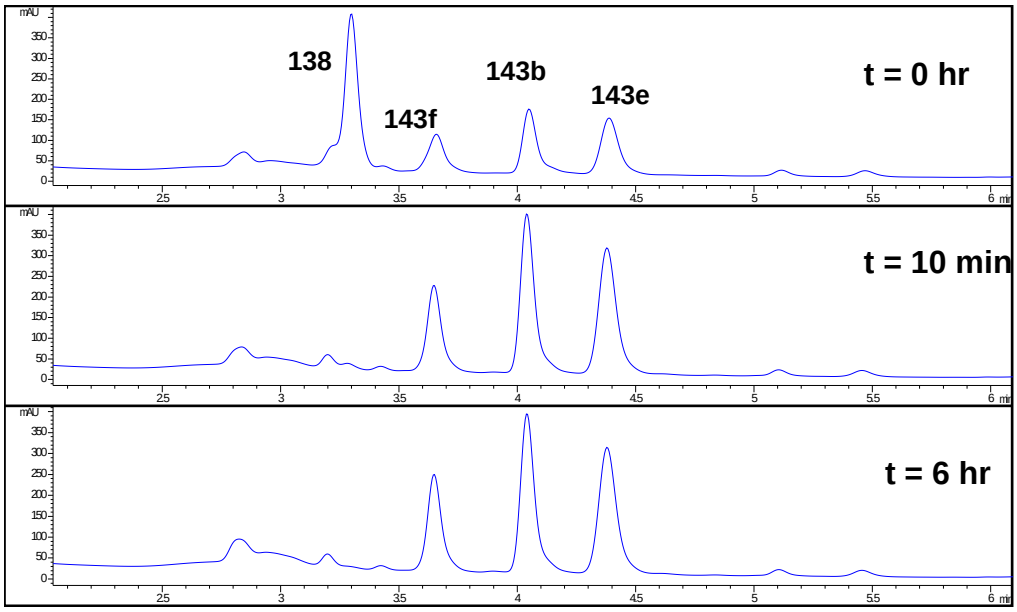


Figure 2.5.a *DCL with 10 mM aniline. Conditions: Aldehyde (1 mM), hydrazides (2 mM each) in NH_4OAc Buffer (100 mM, pH = 6.2) containing 10% DMSO. DCLs were analysed by HPLC at 254 nm.*

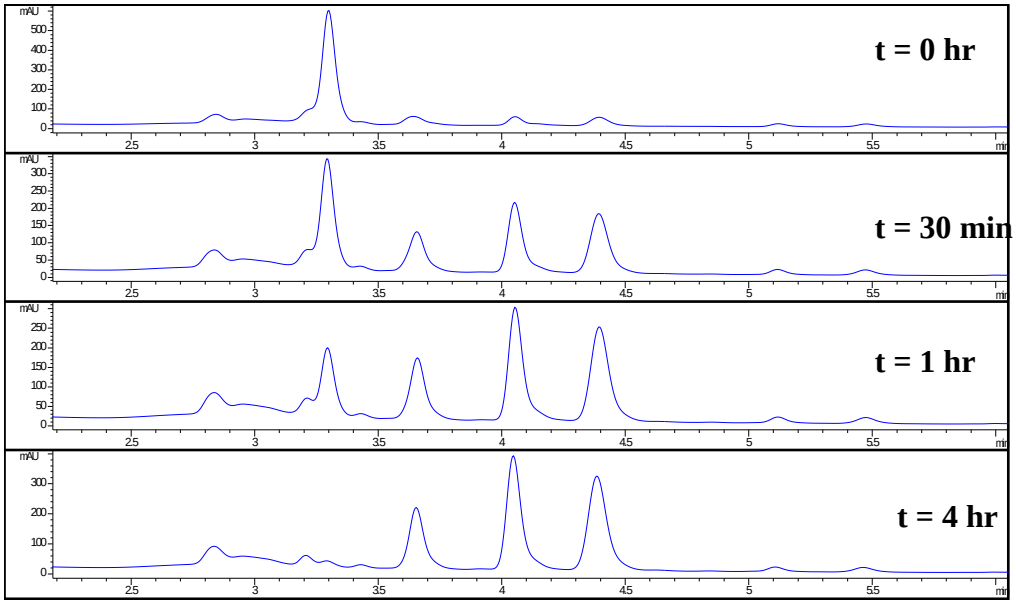


Figure 2.5.b *DCL with 1 mM aniline. Conditions: Aldehyde (1 mM), hydrazides (2 mM each) in NH_4OAc Buffer (100 mM, pH = 6.2) containing 10% DMSO. DCLs were analysed by HPLC at 254 nm.*

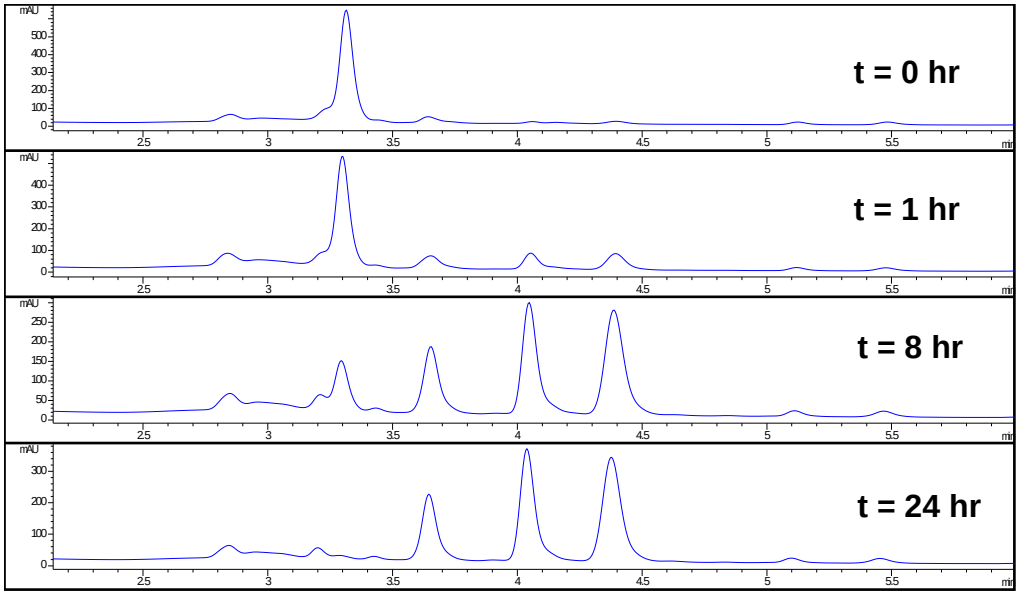


Figure2.5.c *DCL with 0.1 mM aniline. Conditions: Aldehyde (1 mM), hydrazides (2 mM each) in NH_4OAc Buffer (100 mM, pH = 6.2) containing 10% DMSO. DCLs were analysed by HPLC at 254 nm.*

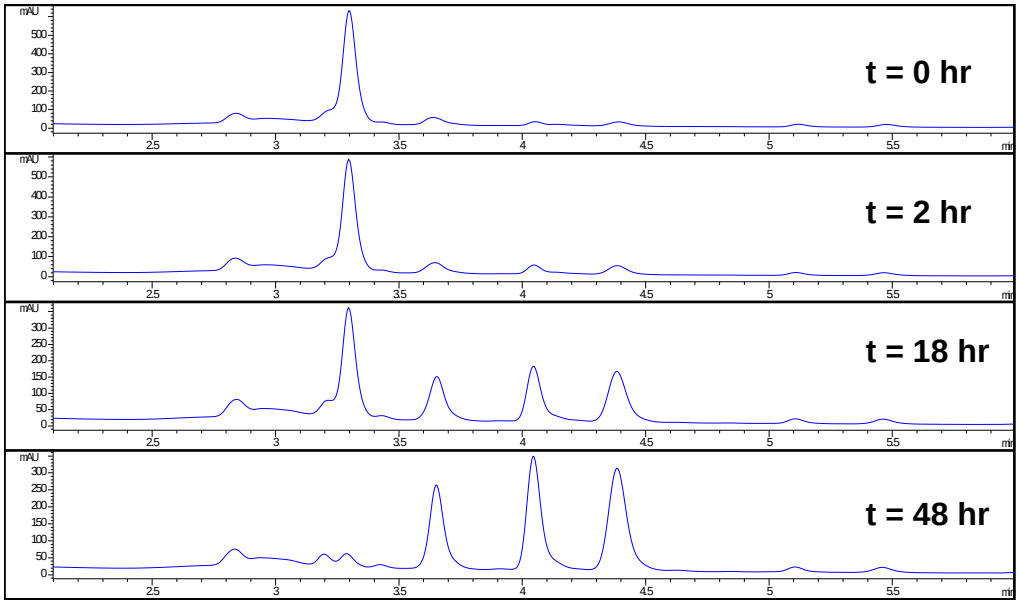


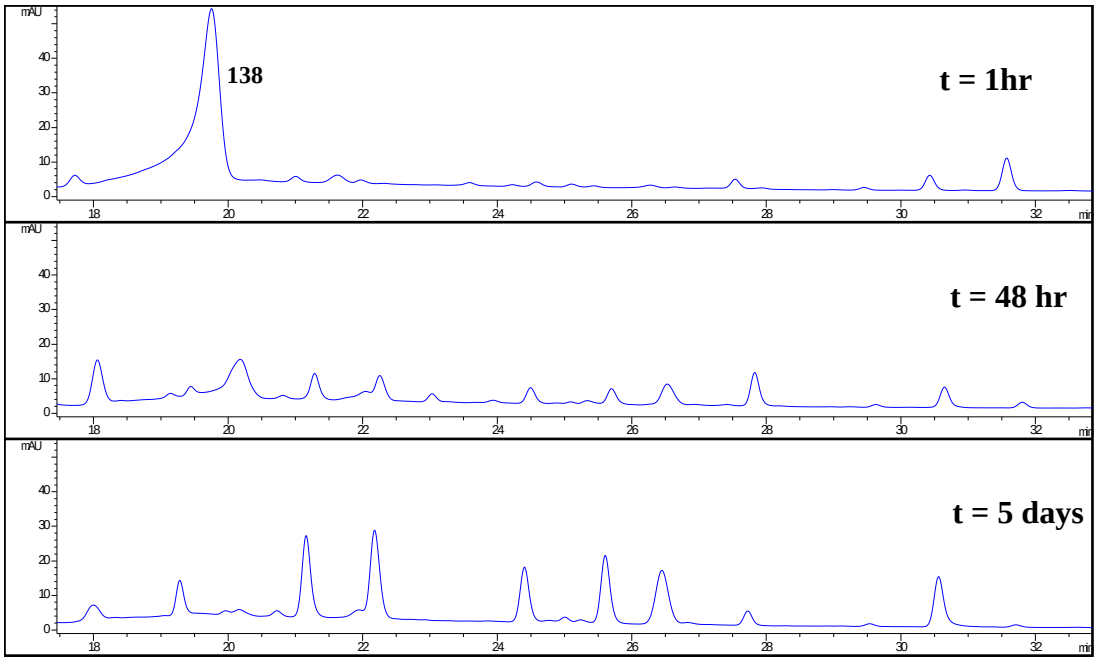
Figure2.5.d *DCL with 0.01 mM aniline. Conditions: Aldehyde (1 mM), hydrazides (2 mM each) in NH_4OAc Buffer (100 mM, pH = 6.2) containing 10% DMSO. DCLs were analysed by HPLC at 254 nm.*

2.4.6 Optimising a 10-membered acylhydrazone library

To understand the effect of aniline on a 10-membered DCL we mixed aldehyde **138** (Scheme 2.8), with 2.5 equivalents of each of the ten aryl hydrazides **141a-h** at room temperature in 15% DMSO. This stoichiometry was chosen to keep the reaction pseudo-first order with respect to the aldehyde so that during the course of the reaction there would not be any shortage of one hydrazide which could introduce a bias in the library.[5] The excess hydrazides used would also limit side reactions between the aldehyde and amino residues of the enzyme. The concentrations of the library constituents (20 μM for the aldehyde and 50 μM for each of the ten hydrazides) were optimised to be within the solubility limits of the hydrazone products formed under the aqueous conditions used since any insoluble compounds can act as thermodynamic traps for the DCL. [32-36]

Equilibration at pH 6.2 was slow, and only two of the ten possible hydrazones could be observed by HPLC after 1 hour (Figure 2.6, DCL A). Despite the presence of excess amounts of the hydrazides there was a significant amount of free aldehyde **138** present throughout the reaction. Equilibrium, as attested by a steady state composition for the library, was attained after five days of incubation. However, a far higher rate of equilibration was observed in the presence of 10 mM aniline as expected from our previous results. The aldehyde **138** could not be detected by HPLC following initial mixing, while a distribution of acyl hydrazones was observed after initial mixing and LC sampling. Complete equilibration of the ten components was observed after just 6 hours (Figure 2.6 DCL B).

DCL A



DCL B

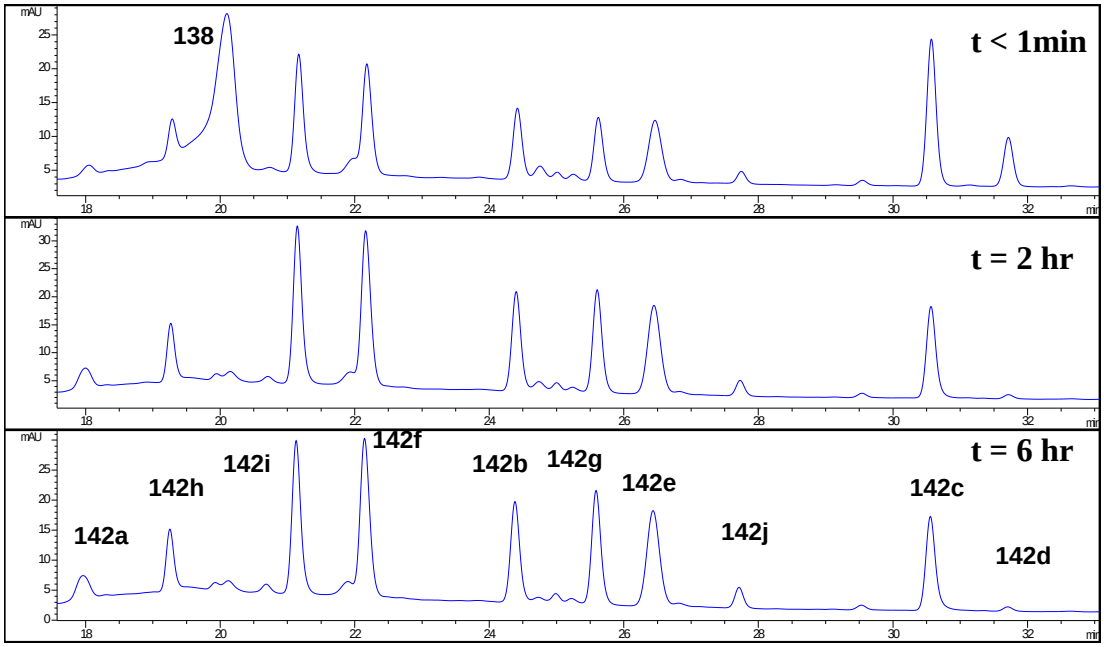


Figure 2.6 *Aniline-catalysed acyl hydrazone DCL. Conditions: Aldehyde (20 μM), hydrazides (50 μM each) in NH_4OAc Buffer (50 mM, pH = 6.2) containing 15% DMSO. Library A is run in the absence of aniline while library B is run in the presence of aniline (10 mM). DCLs analysed by HPLC at 254 nm.*

2.4.7 Reversibility of the DCLs

To ensure adaptive behaviour under the optimised conditions we probed the reversibility of our DCLs. We generated the same library from a different starting composition, hydrazone **142g** plus the nine other hydrazides in the presence of 10 mM aniline. An identical equilibrium distribution was observed within 18 hours, indicating true thermodynamic equilibrium (Figure 2.7). We confirmed the reversibility of the DCL by a second experiment in which an excess of hydrazide **141b** was added to the pre-equilibrated DCL. As expected this resulted in a large amplification of the corresponding acyl hydrazone **142b** (Figure 2.8).

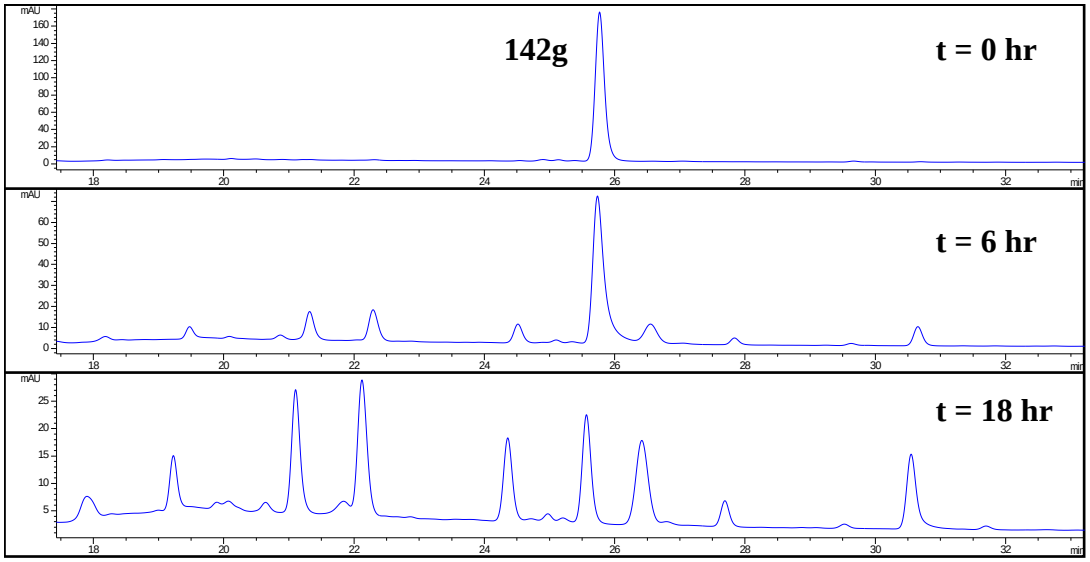


Figure 2.7 *Proof of reversibility of aniline-catalysed acylhydrazone DCLs: constitute DCL from alternative starting point, hydrazone* ***142g****. Conditions: hydrazone* ***142g*** *(20 μM), aniline(10 mM) hydrazides (50 μM each) in NH_4OAc Buffer (50 mM, pH = 6.2) containing 15% DMSO. DCLs analysed by HPLC at 254 nm.*

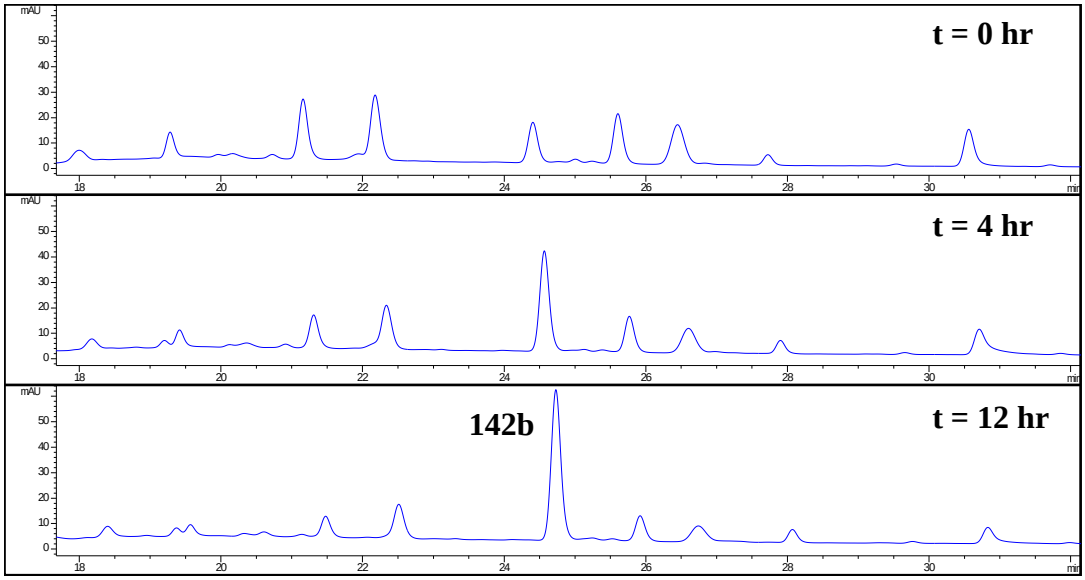
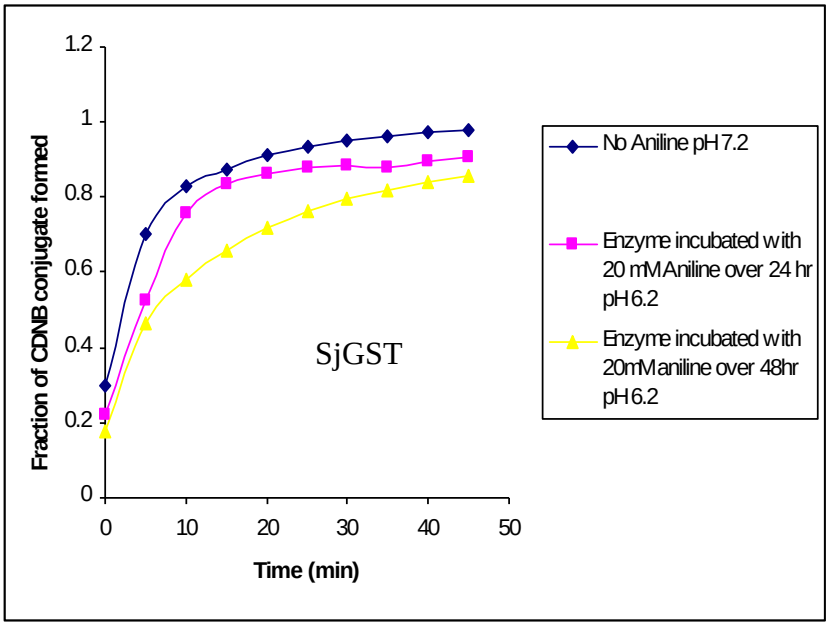


Figure 2.8 *Proof of reversibility of aniline-catalysed acylhydrazone DCLs: addition of excess component (**141b**)(15μL, 10mM, DMSO) to DCL at equilibrium. DCLs analysed by HPLC at 254 nm.*

2.4.8 Protein activity under DCL conditions

Having established nucleophilic catalysis of the DCLs under conditions suitable for biomolecular stability we decided to confirm if our target proteins retained conjugation activity under these optimised conditions. The two recombinant GST isozymes SjGST and hGSTP1-1 were expressed and purified by Anne Caniard from the Campopiano group. We incubated the proteins for two days with 20 mM aniline at pH 6.2 measuring their conjugation activity by a CDNB assay in 24 hour intervals. Both the isoforms retained considerable acitivity under the DCL conditions with 20 mM aniline (Figure 2.9).



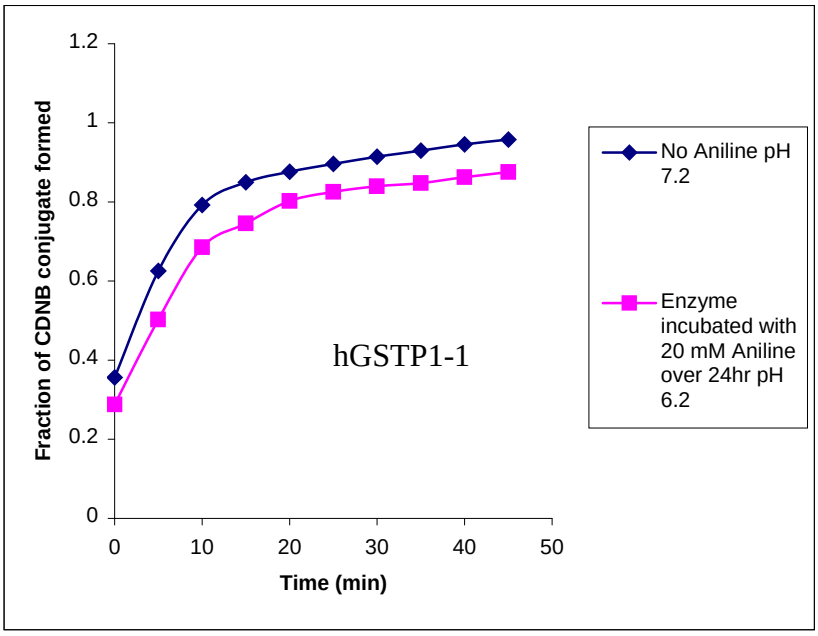


Figure 2.9 *Activity of SjGST and hGSTP1-1 under aniline-catalysed DCL conditions.(Assessed using GSH-CDNB conjugation assay. The samples with aniline had 15% DMSO in ammonium acetate buffer (50 mM, pH 6.2)*

To further confirm the biological compatibility of our optimised aniline-catalysed DCL conditions we synthesised 3-member DCLs (**143b, 143e, 143f**) in the presence of 10 mM GSH and 10 mM lysine respectively. Figure 2.10 shows that the

equilibrium composition of the library is not affected in the presence of large amounts of these biologically relevant nucleophiles.

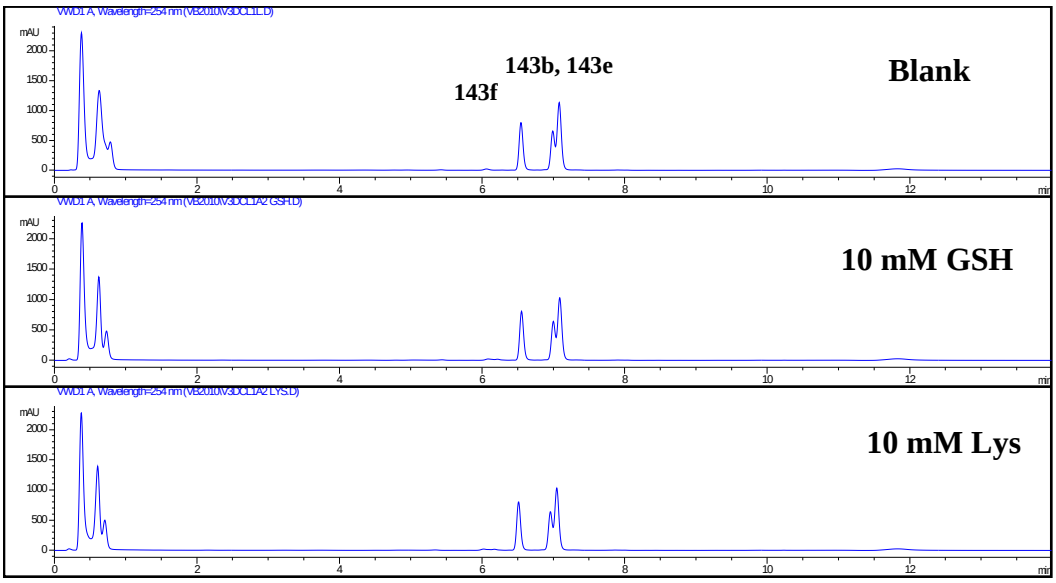


Figure 2.10 *Biological compatibility of acylhydrazone DCLs: 3-membered DCLs equilibrated in the presence of 10 mM GSH and 10 mM Lysine. No change in composition in comparison to the blank DCL*

2.4.9 Protein-templated DCLs

Having optimised biocompatible aniline-catalysed conditions, our next step was to introduce proteins into our DCLs. We synthesized the acyl hydrazone DCL (Scheme 2.8) in the presence of one equivalent each (with respect to the scaffold aldehyde) of the proteins, SjGST and hGST P1-1. These mixtures were equilibrated over 12 hours followed by the addition of base to freeze the library composition. The protein template was removed by centrifugation and the samples analysed by HPLC. The DCL constituted in the presence of SjGST clearly amplified thiophene hydrazone **142g** while hGSTP1-1 amplified t-butylphenyl hydrazone **142c** (Figure 2.11). A control experiment with one equivalent of BSA instead of GST produced no amplification which confirmed that the enzymes were responsible for the selection (Figure 2.12). We re-synthesized the hydrazones in order to assay them but poor solubility prevented the determination of accurate IC_{50} values at the higher concentrations necessary to assay weak binding compounds.

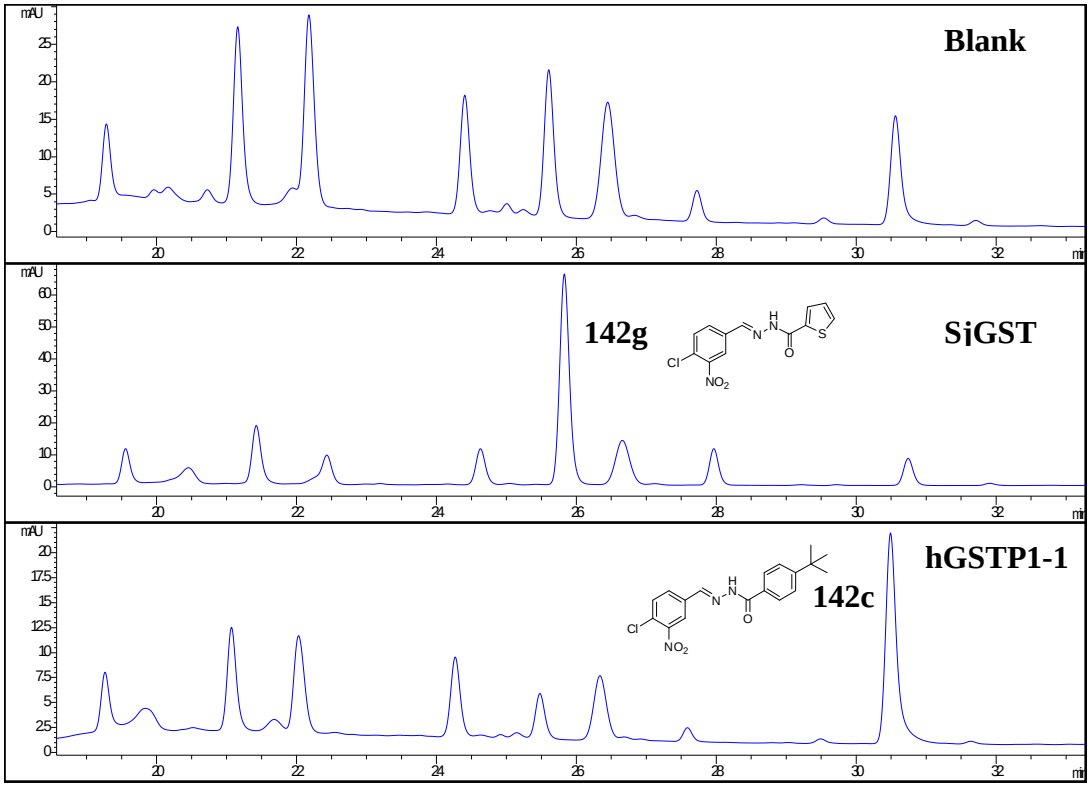


Figure 2.11 *GST-templated DCLs. Conditions: GST (1 equiv), aldehyde (20 μM), hydrazides (50 μM) and aniline (10 mM) in NH_4OAc Buffer (50 mM, pH = 6.2) containing 15% DMSO for 16 h. DCLs analysed by HPLC at 254 nm.*

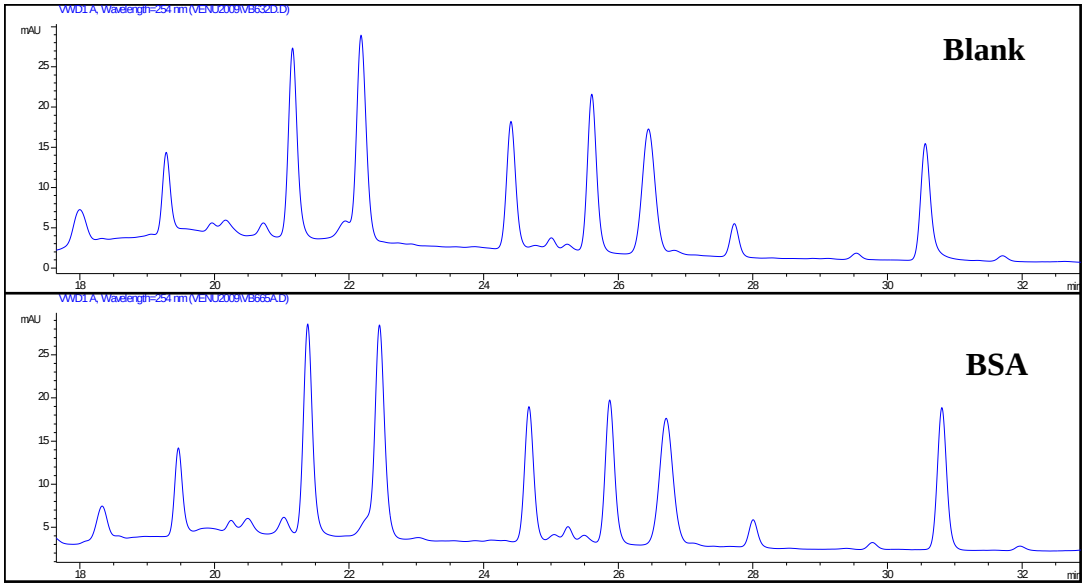
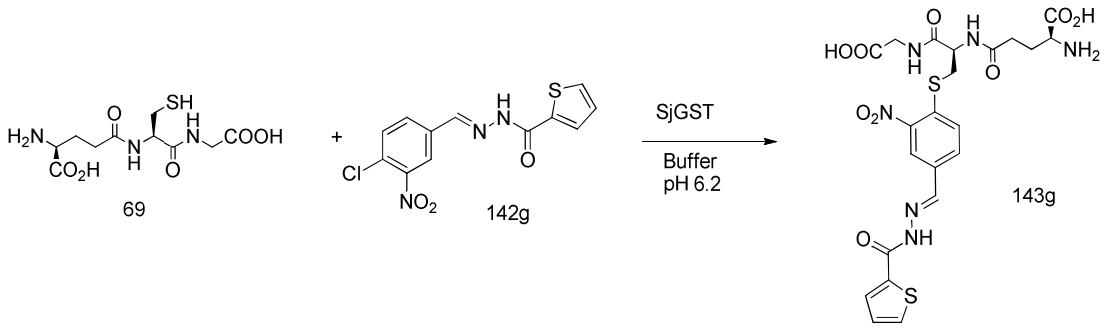
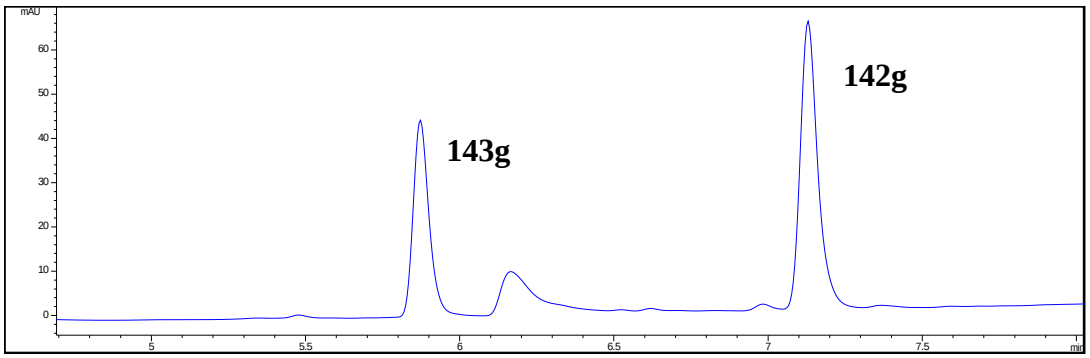


Figure 2.12 *BSA-templated DCLs. Conditions: BSA (1 equiv), aldehyde (20 μM), hydrazides (50 μM) and aniline (10 mM) in NH_4OAc Buffer (50 mM, pH = 6.2) containing 15% DMSO for 16 h. DCLs analysed by HPLC at 254 nm.*

To probe if the amplified components were bound in the target H-region of the enzyme's active site we performed conjugation experiments with GSH. Conjugation of GSH to the aryl chloride group in hydrazones **142g** *via* S_NAr substitution is a slow reaction at pH 6.2, taking several days. In the presence of catalytic amounts of SjGST, however, rapid formation of the S_NAr conjugation adduct for hydrazone **143g** was observed at pH 6.2 (Scheme 2.12). The amplified hydrazone can thus act as a substrate for SjGST and binds in the targeted H-site.



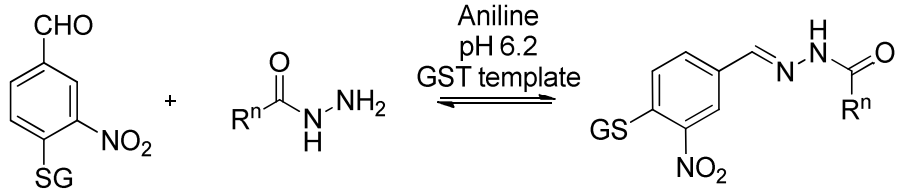


Scheme 2.12 *Demonstration of thiophene hydrazone* ***142g*** *as a GST substrate: Conjugation catalysed by SjGST analysed by LCMS at 254 nm.*

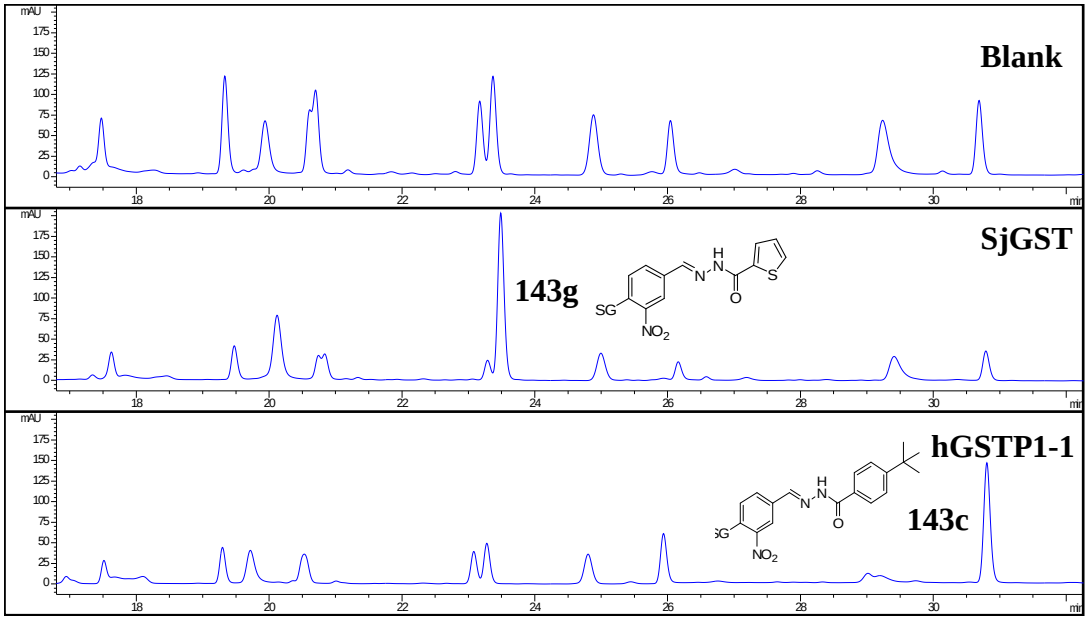
2.4.10 Protein-templated glutathione conjugate DCLs

Anticipating that the GS-conjugate hydrazones would solve the solubility problem and also to capitalize on increased molecular recognition by the protein,[29-31] we synthesized DCLs using GS-conjugated aldehyde **138** and the same ten hydrazides as employed previously (Scheme 2.13). A higher concentration of the aldehyde (50

μM) could be used in this case due to better solubilities of the conjugate hydrazones. Equilibration of the libraries was complete in 6 hours, compared to 4 days in the absence of aniline, and each of the ten acyl hydrazones were identified by LC-MS. Clear amplifications could be observed for DCLs equilibrated in the presence of both GST targets: in each case the same hydrazide fragment was selected as the best binder, thiophene (**143g** for SjGST and t-butylphenyl (**143c**) (Figure 2.13). Both components were amplified to over 300% of their concentrations in the blank DCL, at the expense of nearly all other competing hydrazones (graph in Figure 2.13). Additionally, the anisyl sulfonylhydrazone **143j** underwent *ca.* 100% amplification using hGST P1-1 as the only other positively selected component. Significant reduction in equilibrium concentrations occurred for hydrazones **143b**, **f** and **i** in the case of SjGST and hydrazones **143f, g** and **i** for hGSTP1-1. Whether the protein was present from the beginning of the reaction or was added to a pre-equilibrated library had no bearing on the final equilibrium distribution of the products which indicates true thermodynamic selection.



Scheme 2.13 *Protein-templated conjugate acylhydrazone DCL*



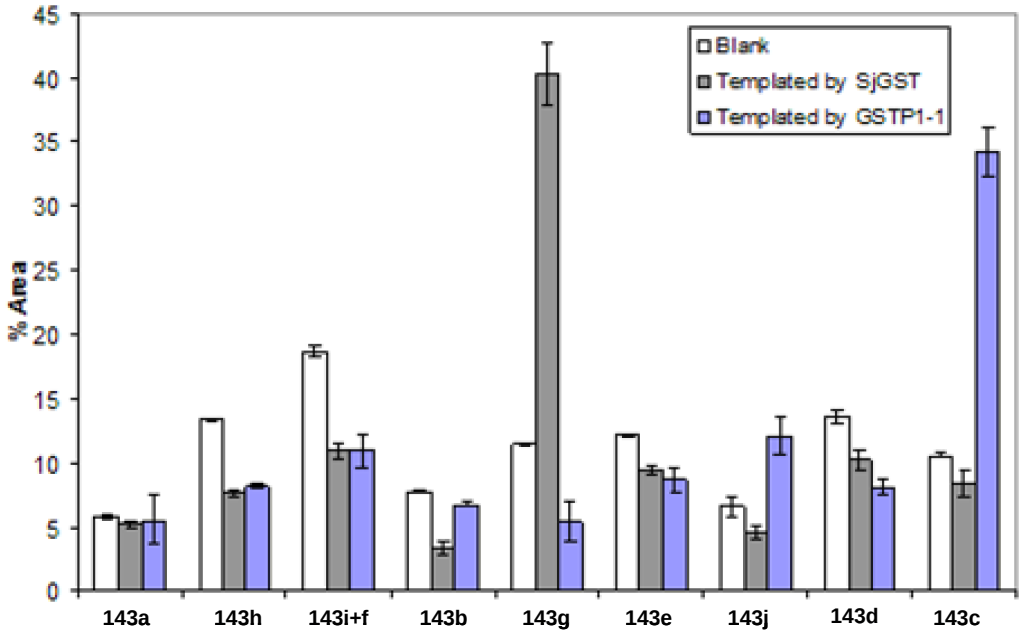


Figure 2.13 *GST-templated DCLs. Conditions: GST (1 equiv), aldehyde (50 μM), hydrazides (50 μM) and aniline (10 mM) in NH_4OAc Buffer (50 mM, pH = 6.2) containing 15% DMSO for 16 h. The graph illustrates changes in DCL component concentration for blank versus SjGST versus hGST P1-1 DCLs. DCLs analysed by HPLC at 254 nm.*

Two different control experiments confirmed that the proteins were responsible for the observed divergent amplification effects. Firstly, synthesizing the DCL in the

presence of one equivalent of BSA produced no amplification (Figure 2.14). Again, both protein-templated DCLs synthesised in the presence of a large excess (100 μM) of the non-selective GST inhibitor ethacrynic acid (**73)** [37-39] yielded no amplification (Figure 2.15). No component amplification is observed in the presence of the inhibitor which indicates that the GST active site is saturated by the ethacrynic acid and cannot influence the DCL equilibrium composition.

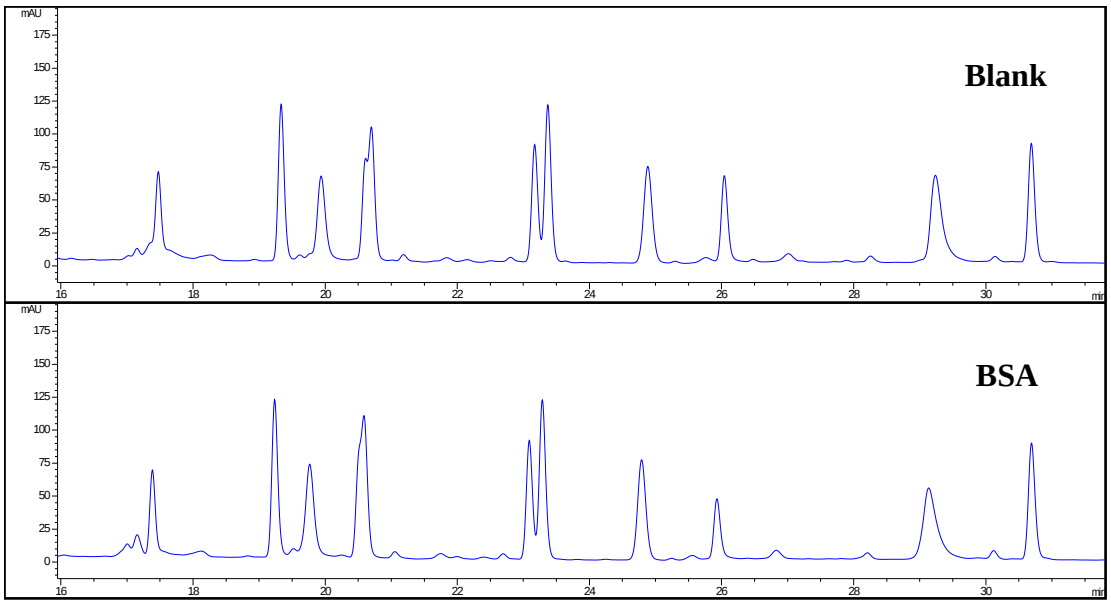
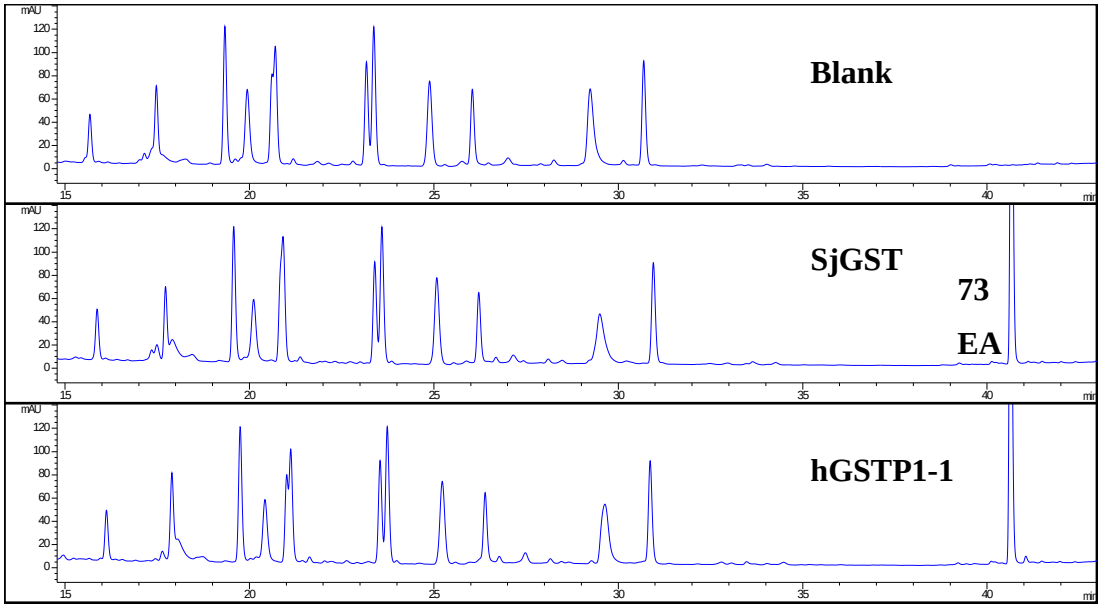
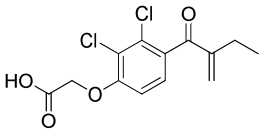


Figure 2.14 *Control Experiment: BSA-templated DCL shows no amplification Conditions: BSA (1 equiv), aldehyde (50 μM), hydrazides (50 μM) and aniline (10 mM) in NH_4OAc Buffer (50 mM, pH = 6.2) containing 15% DMSO for 16 h. DCLs analysed by HPLC at 254 nm.*





*Ethacrynic acid (**73**)*

Figure 2.15 *Control experiment for GSH conjugate DCL in the presence of excess ethacrynic acid (100 μM). DCLs analysed by HPLC at 254 nm.*

2.4.11 Acyl hydrazone DCLs targeting Y7F mutant of SjGST

As a further control experiment and to see if a functionally disabled enzyme would exhibit the same selectivity towards our DCL systems, we choose a Y7F mutant of SjGST as our protein target. Anne Caniard from the Campopiano group expressed and purified the mutant. The protein is catalytically inactive due to the lack of a crucial tyrosine residue which has been replaced with a phenylalanine.[40] The conserved Tyr7 active site residue is known to play a critical role in GSH conjugation for the Sj class of GSTs, stabilising the GSH thiolate anion through H-bonding from the phenol group. We observed essentially zero activity with this mutant in CDNB conjugation versus the wild type SjGST (Figure 2.16, protein activity measured by Anne Caniard). However, SjGST Y7F proved equally effective in controlling DCL composition, showing a clear preference for the same thiophene derivative **143g** as was amplified by the wild type SjGST (Figure 2.16). This

provided crucial evidence that no kinetic selection operates in the DCLs templated by the enzymes.

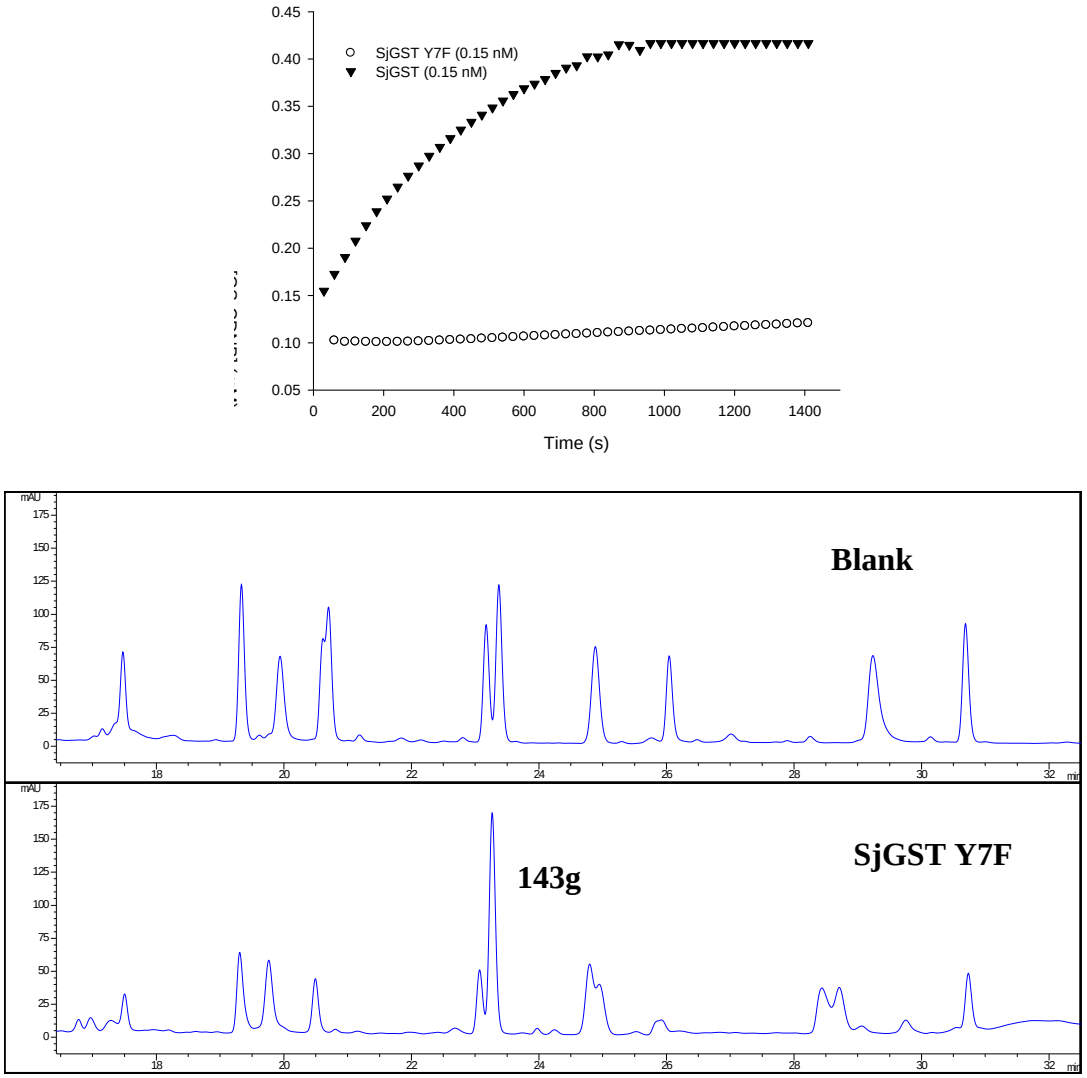


Figure 2.16 *GSH-conjugate DCL templated by SjGST-Y7F mutant amplifying acylhydrazone**143g**. DCL conditions: Y7F SjGST mutant (1 equiv), aldehyde (50 µM), hydrazides (50 µM) and aniline (10 mM) in NH_4OAc Buffer (50 mM, pH = 6.2) containing 15% DMSO for 16 h. The graph shows comparative CDNB assays between SjGST WT and SjGST Y7F. The mutant catalytic activity was reduced to 1.7%.(Assay performed by Anne Caniard)*

2.4.12 Isothermal titration calorimetry

We first confirmed that the amplified ligands are bound to SjGST and hGST P1-1 using isothermal titration calorimetry (ITC). Hydrazones **143a-j** were synthesised

and purified by RP-HPLC (> 95% purity). Isothermal calorimetry was performed on the hydrazones by Anne Caniard in collaboration Prof. Alan Cooper at University of Glasgow. This technique is well established in studying protein-ligand interactions and has unique advantages: simultaneous determination of all associated parameters like the stoichiometry (*N*), the dissociation constant (*K*), the enthalpy (Δ*H*) and the entropy (Δ*S*) changes involved in the binding event.[41-43] These parameters for the amplified ligands **143c** and **143g** bound to SjGST and hGSTP1-1 are given in Table 2.1. Data for scaffold structure **138** is also given for comparison. The enzyme-ligand binding interactions are exothermic with K_d values of compounds **143c** and **143g** an order of magnitude less than the aldehyde **138** which suggests that inhibitor structure has been extended by the hydrazide fragments. The stoichiometry N, gave values around 2 as expected for two ligands per dimer of the protein.

	SjGST-**143c**	SjGST-**143g**	hGST P1-1-**143c**	hGST P1-1-**143g**	SjGST-**138**
K_d (μM)	6.5 ± 0.3	5.8 ± 0.5	3.8 ± 0.4	6.6 ± 1.0	41 ± 3.6
N (stoichiometry)	1.71 ± 0.03	1.71 ± 0.05	2.60 ± 0.06	1.90 ± 0.1	2.02 ± 0.2
Δ*H* (kcal/mol)	-11.4 ± 0.03	-16.9 ± 0.05	-16.9 ± 0.07	-13.4 ± 0.1	-5.0 ± 0.5
Δ*S* (cal/mol)	-14.4	-33.1	-31.8	-21.4	3.2
Δ*G* (kcal/mol)	-7.1	-7.0	-7.4	-6.4	-6.0

W

Table 2.1 *ITC data for Ligand binding to SjGST and hGSTP1-1 (See section 2.6.13 for ITC plots)*

2.4.13 Inhibition assays

In collaboration with A. M. Caniard and D.J Campopiano we studied the inhibitory activity of the hydrazones towards SjGST and hGST P1-1 using the standard CDNB conjugation assay. The hydrazones were synthesized and purity checked by HPLC (> 95%). The IC_{50} values for the different hydrazones are listed in Table 2.2.

Hydrazone	SjGST	hGST P1-1
138	279 ± 23	331 ± 20
143a	24 ± 1	ND
143b	40 ± 2	ND
143c	50 ± 3	57 ± 2
143d	26 ± 1	ND
143e	36 ± 2	ND
143f	61 ± 3	ND
143g	22 ± 1	87 ± 3
143h	34 ± 1	119 ± 8
143i	37 ± 1	84 ± 4
143j	25 ± 1	ND

Table 2.2 *IC_{50} data (μM). Assay conditions: To 360 μl well were added phosphate buffer (255 μl, 0.1 M, pH 6.8), GST (15 μl, 0.15 mg.ml^{-1}) and inhibitor (15 μl, 0.2 to 200 μM). The solution was mixed well and after incubation at 25 °C for 5 min, CDNB (15 μl, 40 mM) and GSH (15 μl, 40 mM) were added quickly and mixed. Absorbance was measured at 340 nm, 25 °C for 5 min.*

All hydrazones exhibited moderate inhibition in the micromolar range for both the isoforms but the values were slightly higher for hGSTP1-1 (59 to 126 μM) compared to SjGST (22 to 63 μM). However, most gratifyingly the DCC selected hydrazones were most active for both the isoforms. Hydrazone **143g** showed and IC_{50} value of 22 μM for SjGST, the lowest among all the library members. Among the 4 conjugates tested against hGSTP1-1 t-butylphenyl hydrazone **143c** had the lowest IC_{50} (57 μM). Thus, by extending the scaffold structure **138** through an adaptive DCC selection process we were able to increase potency by ten-fold for SjGST (279 to 22 μM) relative to the starting anchored aldehyde. A similar six-fold rise in potency was observed for hGSTP1-1 **(**331 to 57 μM).

	hGST P1-1		SjGST	
	CDNB	GSH	CDNB	GSH
K_m (mM)	0.59 ± 0.043	0.38 ± 0.03	4.08 ± 0.80	0.29 ± 0.01
K_i^{143g} (μM)	13.96 ± 0.78	7.19 ± 0.42	12.82 ± 0.68	5.25 ± 0.23
K_i^{143c} (μM)	10.66 ± 0.67	6.61 ± 0.42	18.58 ± 1.05	6.33 ± 0.27
IC_{50}^{143g} (μM)	88 ± 28		22 ± 2	
IC_{50}^{143c} (μM)	59 ± 9		49 ± 10	

Table 2.3 *K_i data (μM)*

The K_i values for the selected hydrazones are given in Table 2.3 and are very close to the K_d values determined by ITC. Interestingly, both the compounds gave slightly higher values when assayed against CDNB in comparison to GSH which binds at the G-site. Both the conjugates showed similar affinities towards the G-sites of the enzymes which is expected for two compounds sharing a common glutathione-tagged nitro-benzene fragment.

2.4.14 Molecular modelling studies

To understand the selectivity of our GST isozymes towards the two hydrazone inhibitors **143c** and **143g**, we carried out a molecular modeling study in collaboration with Dr. Ruth Brenk and Dr. Torsten Luksch (Dundee University). Available GST structures in the protein data bank (PDB) sharing more than 40% sequence identity to the target SjGST crystal structure (1M9A) were analysed and from these 38 structures that contained a bound, GSH-based ligand, were selected. Upon aligning these structures, it was observed that the GSH portions overlaid well, being bound in very similar conformations in the highly conserved G-sites. However, the conjugate

hydrophobic parts of the ligands exhibited great diversity in their bound conformations emphasising the highly promiscuous nature of the enzyme H-site (Figure 2.17a).

A closer examination of the superimposed crystal structures identified the GSH conjugate of 1,2-epoxy-3-(p-nitrophenoxy)propane (EPNP) **144** bound to cGST M1-1 (PDB code 1C72) as the ligand that projected into the H-site with the most similar geometry to the energy-minimized structure of hydrazone **143g** (Figure 2.17b)

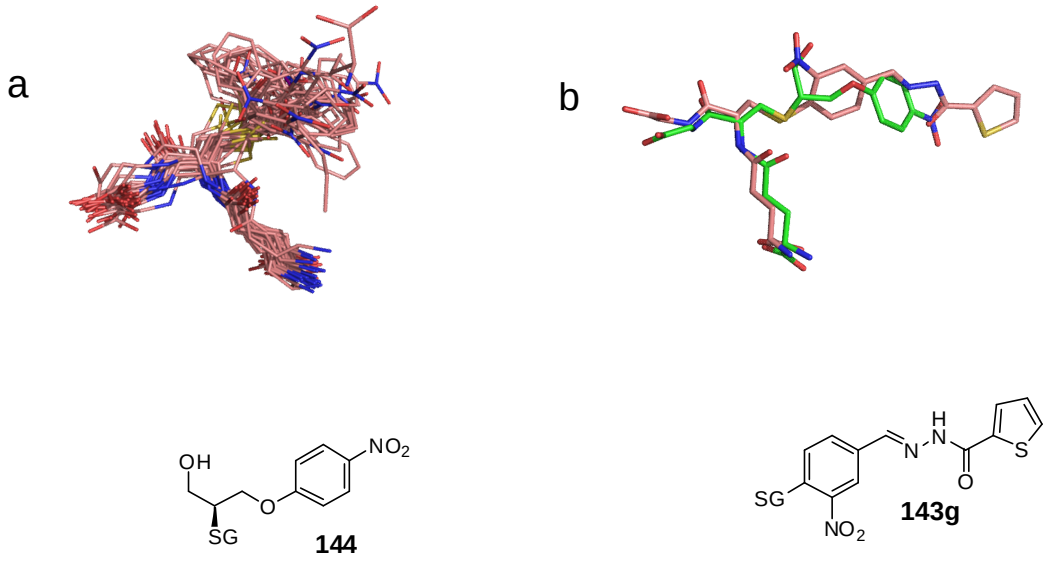


Figure 2.17 a: *Superposition of a selection of GST ligands from the PDB.* ***b:****Conformation of the GST-bound EPNP ligand 144 as found in the crystal structure (PDB code 1c72, green carbon atoms), relative to the energy minimized structure of compound 143g (pink carbon atoms).(Modelling performed by Torsten Luksch and Ruth Brenk, University of Dundee)*

Based on an analysis of the bound ligand **144** a binding model for SjGST was generated with thiophene hydrazone **143g** and for hGST P1-1 with t-butyl hydrazone **143c** (Figure 2.18). The interactions in the generated binding modes for SjGST in complex with **143g** (Figure 2.18a) and hGST P1-1 in complex with **143c** (Figure

2.18b) between the glutathione moiety and the proteins are identical to those reported in previous publications.[38,44] We predict the hydrazone group of **143g** forms hydrogen bonds to R103 and Q204 in SjGST, and equivalent interactions are observed for **143c** in complex with hGST P1-1. Residue V161 in SjGST and I161 in hGST P1-1 make hydrophobic interactions in our models with the ligands **143g** and **143c.**

The thiophene hydrazone fits easily in the SjGST binding pocket with only minor side chain adjustments necessary (RMSD 0.3 Å between model and crystal structure template). The t-butylphenyl group, however, would lead to a steric clash and requires some degree of induced fit in order to bind. Induced fit is also required to accommodate this ligand in the hGST P1-1 pocket, but in this case the binding mode could be stabilized by additional lipophilic interactions of the t-butyl group with Y103, H162 and I161. The amino acid residues at the H-sites of SjGST and hGSTP1-1 are not highly conserved and could play a major role in determining ligand selectivity.

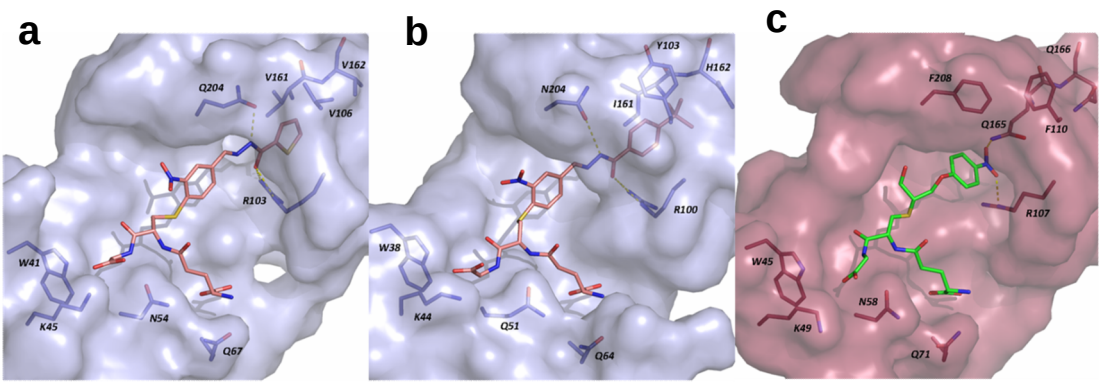


Figure 2.18 *Modelling of amplified ligands at the GST active site. a: Model of* ***143g*** *bound to SjGST. b: Model of* ***143c*** *bound to hGST P1-1.The binding pocket surfaces are shown in light blue and key amino acids as blue sticks. The ligands are represented in salmon pink, with atoms coloured by type. Hydrogen bonds of the conjugated ligand parts are shown as yellow dotted lines. c: The EPNP–cGST M1-1 crystal structure (PDB code 1c72). The binding pocket surface is shown in raspberry pink and key amino acids as red sticks. The ligand is represented in green, with atoms coloured by type.Hydrogen bonds of the conjugated ligand parts are shown as yellow dotted lines.(Modelling performed by Torsten Luksch and Ruth Brenk, University of Dundee)*

2.5 Conclusions

We have established that adaptive DCls based on reversible synthesis of acyl hydrazones can be compatible with protein targets by using aniline as a nucleophilic catalyst. The pH (= 6.2) used was optimised as a fine balance between facile reversibility of the acylhydrazone reaction and protein stability. We then used these DCLs as a tool to explore the promiscuous H-sites of various GST isoforms. The two GST isoforms showed divergent amplification effects with the selected hydrazones showing increased inhibitory activity of over one order of magnitude from the starting GSH-tagged nitro-benzaldehyde **138**, thus validating the approach in the context of protein–ligand discovery. The library size was kept deliberately small in order to fully characterise equilibrium distributions and quantify amplifications by separation. Isoform-specific ligands could still be identified from such small libraries which indicate the inherent strength of this evolutionary method. The fine structural features that define the active sites of the enzymes that discriminate between the selected hydrazones are still unclear; a deeper understanding would require structural determination of the complexes of GST: GS-hydrazone conjugates.

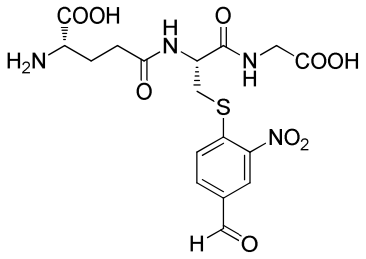
2.6 Experimental

2.6.1 General Methods

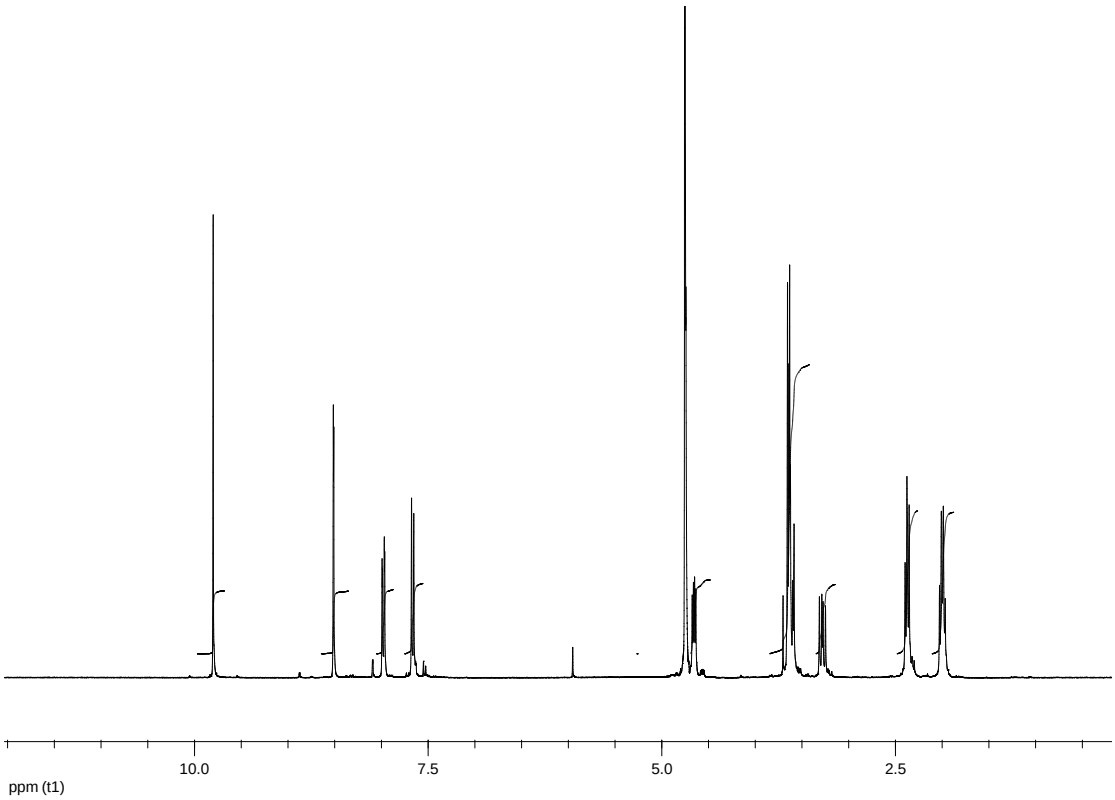
Melting points are uncorrected. ^{1}H and ^{13}C NMR spectra were recorded on a Brüker dpx360 (360 MHz) instrument and calibrated to residual solvent peaks: proton ($CDCl_3$ 7.26 ppm) and carbon ($CDCl_3$ 77.0 ppm). The ^{1}H data is presented as follows: chemical shift (in ppm on the δ scale), multiplicity (s=singlet, d=doublet, t=triplet, q=quartet, m=multiplet), the coupling constant (J, in Hertz) and integration. The ^{13}C data is reported as the ppm on the δ scale followed by the interpretation. Electrospray and electron impact high resolution mass spectrometry was performed by the EPSRC National Mass Spectrometry Service Centre, Swansea, using a Finnigan MAT 900 XLT double focusing mass spectrometer. The data is recorded as the ionisation method followed by the calculated and measured masses. TLC was

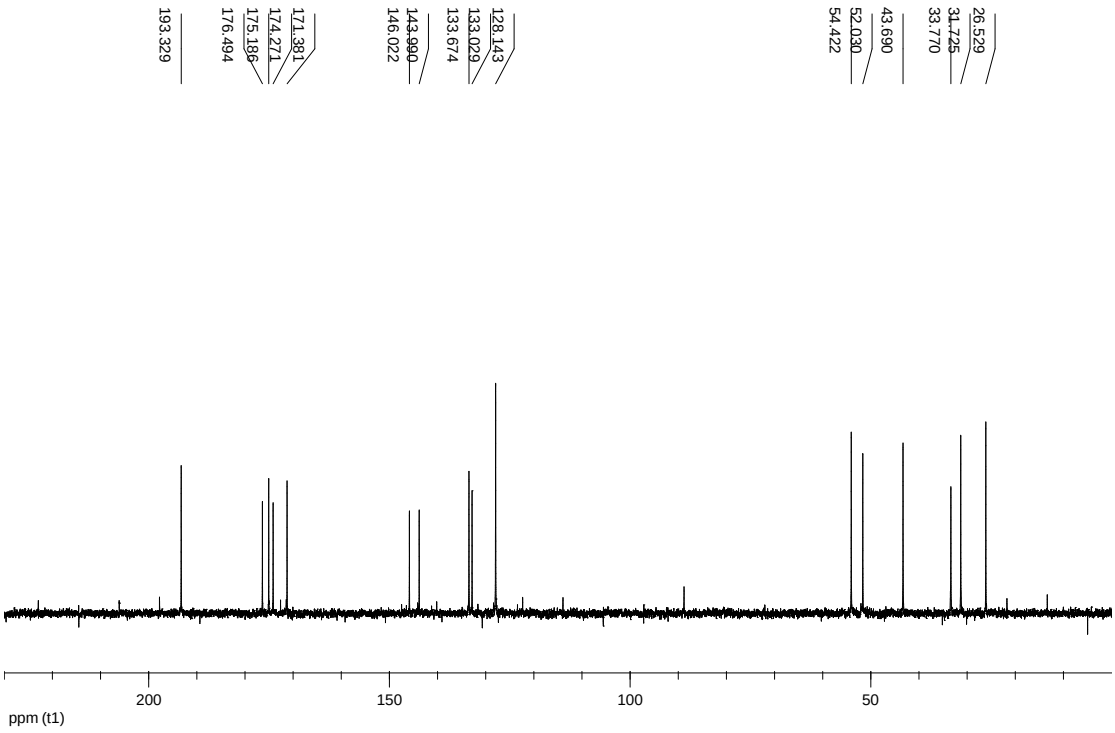
performed on Merck 60F254 silica plates and visualised by UV light and/or anisaldehyde or potassium permanganate stains. All chemicals were purchased from a chemical supplier and used as received. Analytical and preparative HPLC were performed on instruments from Agilent Technologies, UK.

2.6.2 Synthesis of glutathione conjugate aldehyde 138



In a 50 mL round bottom flask, 4-chloro-3-nitrobenzaldehyde **137** (500 mg, 2.69 mmol) and glutathione (994 mg, 3.23 mmol) were dissolved in a mixture of water/MeCN (15 mL, 1:1). Sodium carbonate (343 mg, 3.23 mmol) was then added and the reaction mixture stirred at room temperature for 48 hours. The yellow coloured solution was washed with diethyl ether (2 × 15 mL) and the aqueous phase was concentrated to approximately 5 mL followed by purification by RP-HPLC. The column fractions were pooled together and lyophilized to give the pure compound as a yellow fluffy solid. (66% yield). ^{1}H NMR (d^{6}-DMSO, 360 MHz) δ_H 9.81 (s, 1H), 8.52 (d, *J* = 1.7 Hz, 1H), 7.99 (dd, *J* = 8.5, 1.7 Hz, 1H), 7.67 (d, *J* = 8.5 Hz, 1H), 4.67 (dd, *J* = 9.2, 4.7 Hz, 1H), 3.67-3.60 (m,4H) 3.30 (dd, *J* = 14.0, 9.30 Hz, 1H) 2.49-2.37 (m, 1H) 2.02 (dd, *J* = 14.6, 7.07 Hz, 1H).^{13}CNMR (d^{6}-DMSO, 90 MHz), 193.3, 176.5, 175.2, 174.3, 171.4, 146.0, 144.0, 133.7, 133.0, 128.1 (2 signals overlapping), 54.4, 52.0, 43.7, 33.8, 31.7, 26.5. HRMS calcd for [$C_{17}H_{21}N_4O_9S$] (MH^+) requires m/z 457.1024; found 457.1019 (ESI^+). M.p. 200 ºC (darkens), > 280 ºC (decomp.).


10.0
7.5
5.0
2.5
ppm (t1)


193.329
176.494
175.186
174.271
171.381
146.022
143.990
133.674
133.029
128.143
54.422
52.030
43.690
33.770
31.725
26.529
200
150
100
50
ppm (t1)

2.6.3 General method for the synthesis of hydrazones

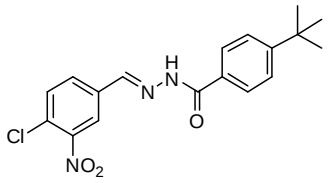
In a 25 mL round bottom flask the aldehyde **137**(1 eq) and the corresponding hydrazide (1.1 eq) were dissolved in 10 ml of a mixture of 7:3 water/acetonitrile followed by the addition of a few drops of glacial acetic acid. The resulting yellow precipitate was filtered off, washed many times with cold ether and cold water and then lyophilized to give the pure hydrazone.

2.6.4 General method for the synthesis of glutathione conjugated hydrazones

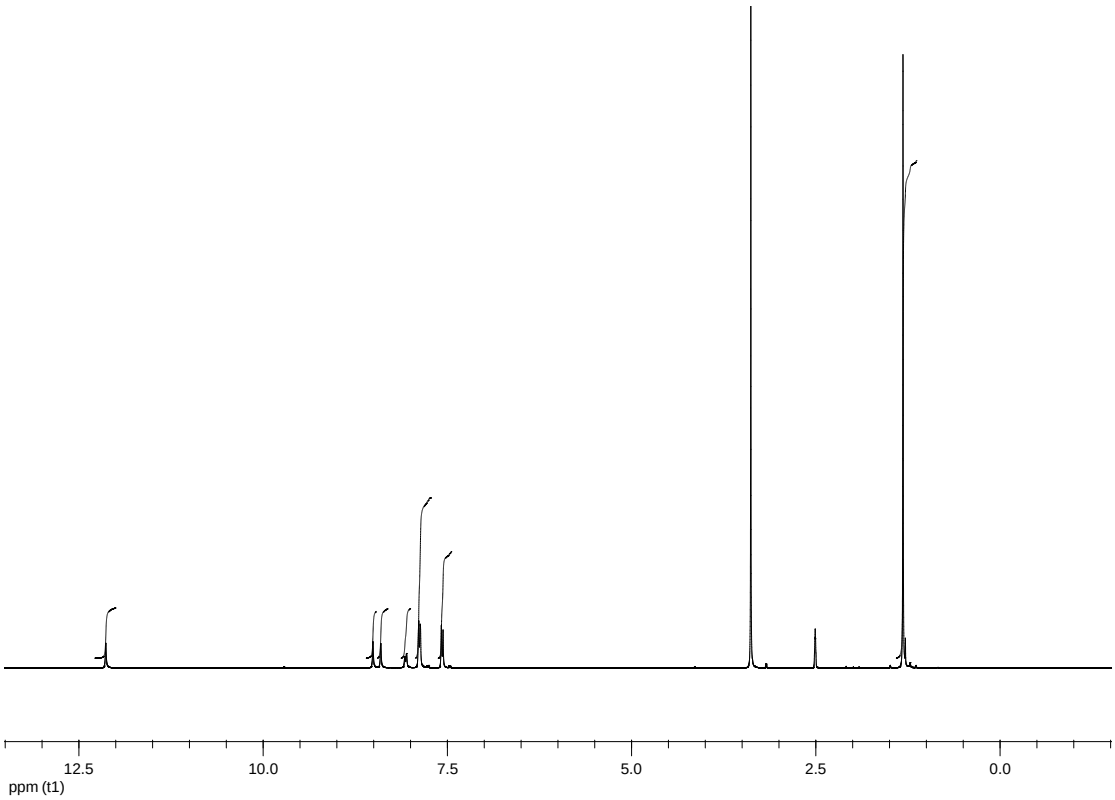
In a 25 mL round bottom flask the glutathione conjugated aldehyde **138** (1 eq) and the corresponding hydrazide (1.1 eq) were dissolved in 10 mL of a mixture of 7:3 water/acetonitrile followed by the addition of a few drops of glacial acetic acid. The resulting yellow precipitate was filtered off, washed many times with cold ether and then lyophilized to give the pure hydrazone as solid.

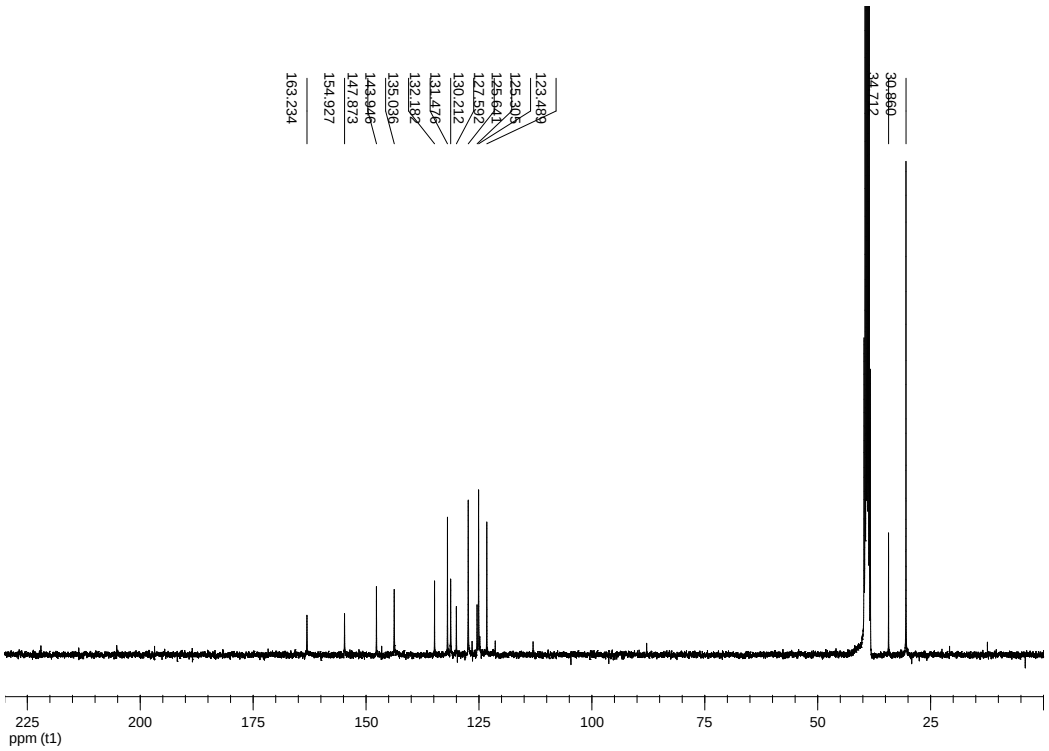
Representative hydrazone data

Hydrazone **142c**:

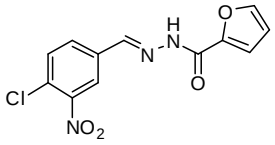


White solid (81%yield). ^{1}H NMR δ_H 12.13 (s, 1H), 8.52 (s, 1H) 8.41 (s, 1H), 8.08 (d, *J* = 8.2 Hz, 1H), 7.89 (m, 3H), 7.58 (d, *J* = 8.3 Hz, 2H), 1.34 (s, 9H) ^{13}CNMR (d^6-DMSO, 90 MHz), 163.2, 155.0, 147.9, 144.0, 135.0, 132.2, 131.5, 130.2, 127.5, 125.6, 125.3, 123.5, 34.7, 30.9. HRMS calcd for [$C_{18}H_{19}ClN_3O_3$] (MH^+) requires m/z 360.1060; found 360.1063 (ESI^+). M.p 268-270 ºC (EtOH).

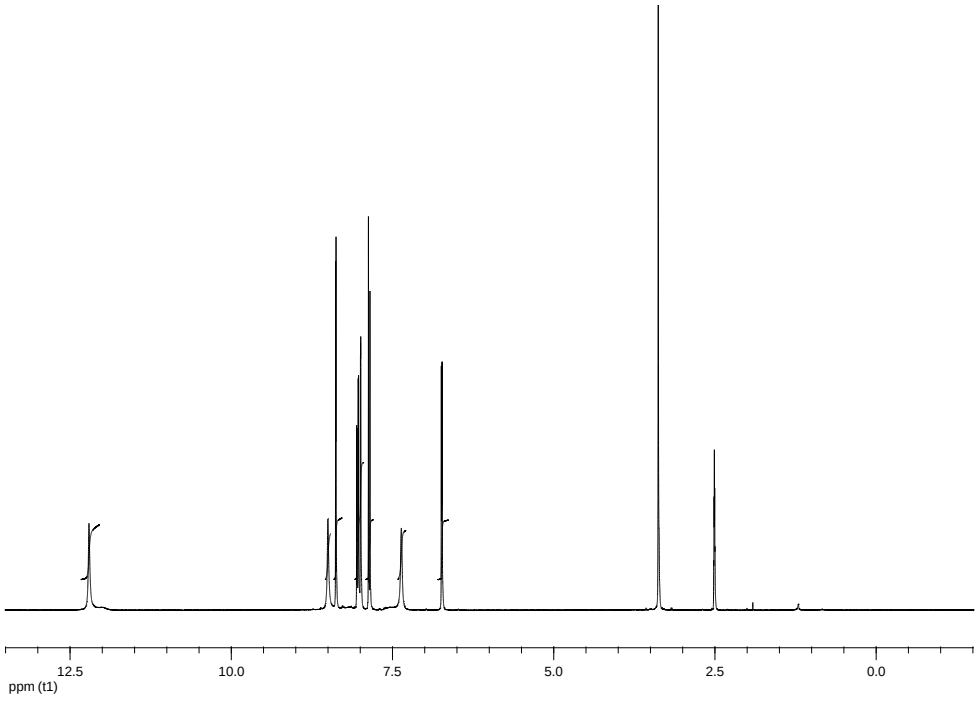

12.5
10.0
7.5
5.0
2.5
0.0
ppm (t1)

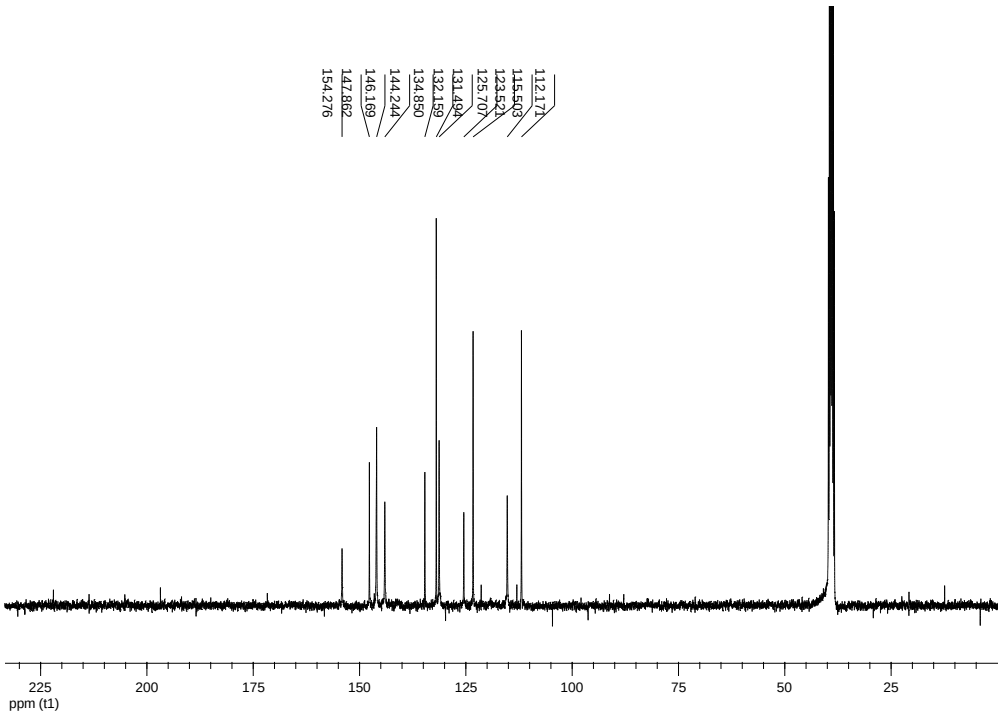

163.234
154.927
147.873
143.946
135.036
132.182
131.476
130.212
127.592
125.641
125.305
123.489
34.712
30.860
225
200
175
150
125
100
75
50
25
ppm (t1)

Hydrazone **142f**:



White solid (76% yield). ^{1}H NMR δ_H 12.21 (s, 1H), 8.51 (s, 1H), 8.39 (d, *J* = 1.9 Hz, 1H), 8.05 (dd, *J* = 8.4, 2.02 Hz, 1H), 8.01 (d, *J* = 0.9 Hz, 1H), 7.87 (d, *J* = 8.4 Hz, 1H), 7.38 (s, 1H), 6.75 (dd, *J* = 3.5, 1.7 Hz, 1H). ^{13}CNMR (d^6-DMSO, 90 MHz), 154.3, 147.9, 146.2, 144.2, 134.9, 132.2, 131.5, 125.7, 123.5, 115.5, 112.2. HRMS calcd for [$C_{12}H_9ClN_3O_4$] (MH^+) requires m/z 294.0231; found 294.0233 (ESI+). M.p 193-194 ºC (EtOH).





2.6.5 GST-catalysed glutathione conjugation of aldehydes 137 and 139

Reduced glutathione (20 µL, 10mM aqueous), aldehyde **137** (20 µL, 10mM, DMSO) and SjGST (10 µL, 180 µM in KPhos buffer 0.1 M, pH 6.8) were added to ammonium acetate buffer (950 µL 50mM , pH 7.2) at room temperature. After 2 hours the protein was removed by filtration using a 10000 MWCO filter (Vivaspin) and the resulting solution analysed by LC-MS. The experiment was repeated with aldehyde **139**.

2.6.6 Aniline catalysis of acylhydrazone formation: effect of substituted anilines

Aldehyde **138** (10 µL, 50 mM aqueous) and hydrazide **141b** (10 µL, 50 mM, DMSO) were added to a mixture of DMSO (90 µL) and ammonium acetate buffer (890 µL 0.1M, pH 6.2). The fraction of the hydrazone (**143b**) formed at room temperature was monitored by HPLC over 8 hours. (HPLC conditions: Column: Luna 5 µ C18(2) 30 mm × 4.6 mm, flow rate 1mL min-1, wavelength 254 nm, temperature 30 ^{o}C, gradient H_2O / MeCN (0.01% TFA) from 95% to 30% over 4 min). The concentrations and fractions were calculated from the integrals of the HPLC signals

at 254nm. The reaction was then performed in the presence of substituted anilines and imidazole respectively (10 mM).

2.6.7 Effect of aniline concentration on DCL equilibration: Single hydrazone formation

Aldehyde **138** (20 µL, 50 mM aqueous) and hydrazide **141b** (20 µL, 50 mM, DMSO) were added to a mixture of DMSO (80 µL) and ammonium acetate buffer (880 µL 0.1M, pH 6.2). The fraction of the hydrazone (**143b**) formed at room temperature was monitored by HPLC over 8 hours. (HPLC conditions: Column: Luna 5 µ C18(2) 30 mm × 4.6 mm, flow rate 1mL min-1, wavelength 254 nm, temperature 30 ^{o}C, gradient H_2O / MeCN (0.01% TFA) from 95% to 30% over 4 min). The reaction was then performed under increasing concentrations of aniline (0.01-10 mM). The concentrations and fractions were calculated from the integrals of the HPLC signals at 254nm.

2.6.8 Effect of aniline concentration on DCL equilibration: 3 membered DCL

Hydrazides **141b**, **141e** and **141f** (3 × 20 µL, 0.1 M, DMSO), aldehyde **138** (20 µL, 50 mM, aqueous) and aniline (20 µL, 500 mM, DMSO) were added to a mixture of DMSO (10 µL) and ammonium acetate buffer (880 µL, 0.1M, pH 6.2). The DCL with an aniline concentration of 10mM was allowed to stand at room temperature with occasional shaking and monitored periodically by HPLC to establish the composition until the relative concentrations of the hydrazones became constant. (HPLC conditions: Column: Luna 5 µ C18(2) 30 mm × 4.6 mm, flow rate = 1 mL min-1, wavelength 254 nm, temperature 30 ^{o}C, gradient H_2O / MeCN (0.01% TFA) from 95% to 80% over 1.5 min then to 60% over 3 min and eventually to 30% over 1.5 min). The DCL was then re-generated and analysed over time in the presence of 1 mM, 0.1 mM and 0.01 mM aniline concentration respectively.

2.6.9 Aniline catalysis of ten-membered hydrazone library

The ten hydrazides **141a-j** (10 × 5 µL, 10 mM, DMSO), aldehyde **137** (2 µL, 10 mM, DMSO) and aniline (10 µL, 1 M, DMSO) were added to a mixture of DMSO

(93 µL) and ammonium acetate buffer (845 µL, 50 mM, pH 6.2). The DCL was allowed to stand at room temperature with occasional shaking and monitored periodically by HPLC to establish the blank composition till the relative populations of the hydrazones became constant. The pH of all samples was raised to 8 by the addition of NaOH (15 µL, 1M, aqueous). LC-MS verified that each of the expected hydrazones was present in the DCL (HPLC conditions: Column: Luna 5 µ C18(2) 30 mm × 4.6 mm and Luna 5 µ C18(2) 50 mm × 4.6 mm in sequence, flow rate : 1mL min^{-1}, wavelength 254nm, temperature 23°C, gradient H_2O/ MeCN (0.01% TFA) from 95% to 80% over 6 min then to 45% over 30 min and eventually to 5% over 5 min). The DCL was then re-synthesised in the absence of aniline and the HPLC traces at different time intervals were compared.

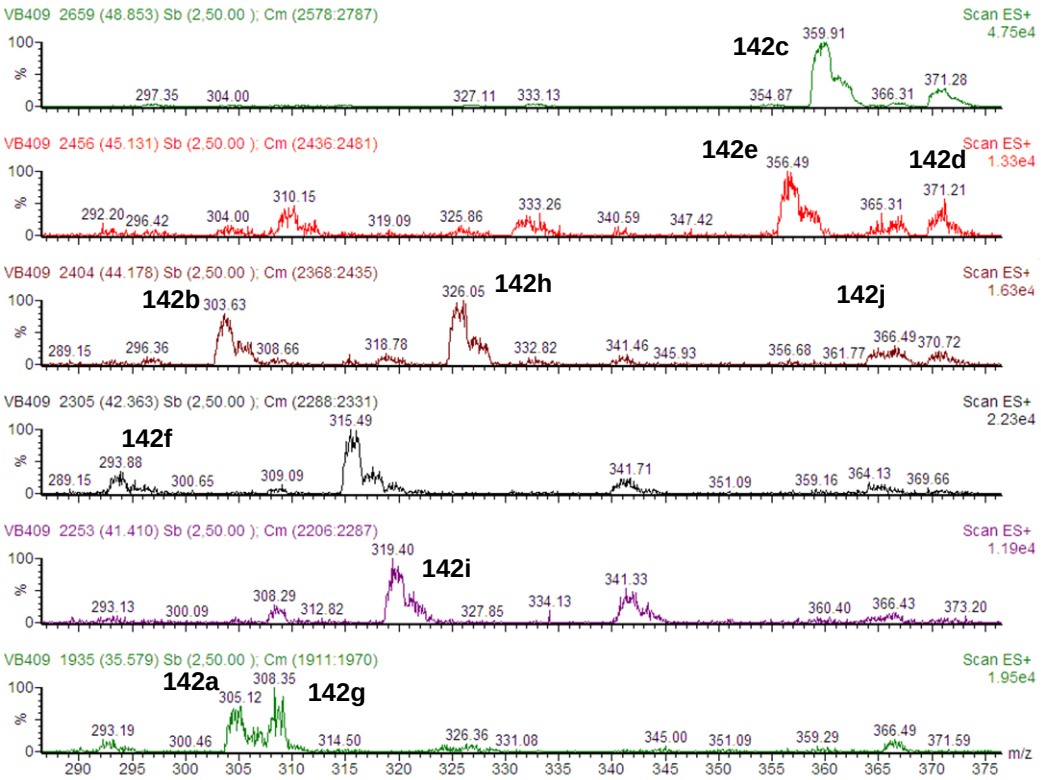


Figure 2.19 *MS data for DCL based on aldehyde* ***137****, indicating that all the expected ten acylhydrazones are present*

2.6.10 Protein activity under DCL conditions

Reduced glutathione (20 µL, 10mM aqueous), CDNB (20 µL, 10mM, DMSO) and SjGST (10 µL, 180 µM in KPhos buffer 0.1 M, pH 6.8) were added to ammonium acetate buffer (950 µL 50mM , pH 7.2). The fraction of glutathione conjugate formed was monitored in every five minute intervals by HPLC over 50 min. (HPLC conditions: Column: Luna 5 µ C18(2) 30 mm x 4.6 mm, flow rate 1mL min^{-1}, wavelength 254nm, temperature 23^{o}C, gradient H_2O/ MeCN (0.01% TFA) from 95% to 20% over 4 min)

SjGST (10 µL , 180 µM in KPhos buffer 0.1 M, pH 6.8) was added to a mixture of DMSO (110 µL) ,ammonium acetate buffer (840 µL 50mM, pH 6.2) and aniline (20 µL, 1M, DMSO) Two samples of these mixtures were incubated at 25^{o}C over 24 hours and 48 hours respectively. To both the samples were added reduced glutathione (20 µL, 10mM aqueous) and CDNB (20 µL, 10mM, DMSO) and the fraction of glutathione conjugate formed was measured for five minute intervals.

2.6.11 Protein-templated DCLs

SjGST (111 µL, 180 µM, in potassium phosphate buffer 0.1 M, pH 6.8), the ten hydrazides **141a-j** (10 × 5 µL, 10 mM, DMSO), the aldehyde **137** (2 µL, 10 mM, DMSO) and aniline (10 µL, 1 M, DMSO) were added to a mixture of DMSO (93 µL) and ammonium acetate buffer (734 µL, 50 mM, pH 6.2). The DCL was allowed to stand at room temperature with occasional shaking over 12 hours. The pH of the sample was raised to 8 by the addition of NaOH (15 µL, 1M, aqueous) and the protein was removed by ultrafiltration using a 10000 MWCO filter (Vivaspin). HPLC analysis was performed and the traces were compared with the blank composition. (HPLC conditions: Column: Luna 5 µ C18(2) 30 mm × 4.6 mm and Luna 5 µ C18(2) 50 mm × 4.6 mm in sequence, flow rate : 1mL min^{-1}, wavelength 254nm, temperature 23^{o}C, gradient H_2O/ MeCN (0.01% TFA) from 95% to 80% over 6 min then to 45% over 30 min and eventually to 5% over 5 min).

The DCL composition was identical regardless of whether the SjGST was present from the beginning, or added after pre-equilibration, but equilibration took more than 24 hours in the latter case.

For the hGST P1-1 templated library, the ten hydrazides **141a - j** (10 × 5 µL, 10 mM, DMSO), the aldehyde **137** (2 µL, 10 mM, DMSO), aniline (10 µL, 1 M, DMSO) and hGST P1-1 (100 µL, 200 µM, in potassium phosphate buffer 0.1 M, pH 6.8) were added to a mixture of DMSO (93 µL) and ammonium acetate buffer (734 µL, 50 mM, pH 6.2). After equilibration over 12 hours the DCL was analysed using HPLC. Control experiments were performed using the same equivalents of BSA in place of GST.

2.6.12 Protein-templated glutathione conjugate DCLs

To establish the blank DCL composition the ten hydrazides **141a - j** (10 × 5 µL, 10mM, DMSO), aldehyde **138** (5 µL, 10mM, aqueous) and aniline (10 µL, 1M, DMSO) were added to a mixture of DMSO (96 µL) and ammonium acetate buffer (839 µL, 50 mM, pH 6.2). The DCL was allowed to stand at room temperature with occasional shaking and monitored periodically by HPLC to establish the blank composition till the relative populations of the hydrazones became constant. The pH of all samples was raised to 8 by the addition of NaOH (15 µL, 1M, aqueous). LC-MS verified that each of the expected hydrazones was present in the DCL. (HPLC conditions: Column: Luna 5 µ C18(2) 50 mm × 4.6 mm and Luna 5 µ C18(2) 250 mm × 4.6 mm in sequence, flow rate : 1mL min^{-1}, wavelength 254nm, temperature 23^{o}C, gradient H_2O/ MeCN (0.01% TFA) from 95% to 5% over 40 min).

For resynthesizing the DCL in the presence of the protein SjGST (278 µL, 180 µM, in potassium phosphate buffer 0.1 M, pH 6.8), the ten hydrazides **141a - j** (10 × 5 µL, 10mM, DMSO), aldehyde **138** (5 µL, 10 mM, DMSO) and aniline (10 µL, 1 M, DMSO) were added to a mixture of DMSO (96 µL) and ammonium acetate buffer (561 µL, 50 mM, pH 6.2). The DCL templated by hGST P1-1 was synthesized by adding the ten hydrazides **141a - j** (10 × 5 µL, 10mM, DMSO), the aldehyde **138** (5 µL, 10 mM, DMSO), aniline (10 µL, 1 M, DMSO) and hGST P1-1 (250 µL, 200

µM, in KPhos buffer 0.1 M, pH 6.8) to a mixture of DMSO (96 µL) and ammonium acetate buffer (589 µL, 50 mM, pH 6.2). The DCLs were allowed to stand at room temperature over 12 hours after which the pH was raised to 8 by the addition of NaOH (15 µL, 1M). The protein was filtered off using a centrifuge filter of MWCO 10000 followed by analysis of the filtrate by HPLC using similar conditions to those listed above. Control experiments were performed using the same equivalents of BSA in place of GST. For the control experiments with ethacrynic acid **73**, a large excess (100µM) of the inhibitor was used. A similar DCL was equilibrated in the presence of the YF7 mutant of SjGST and analysed by HPLC.

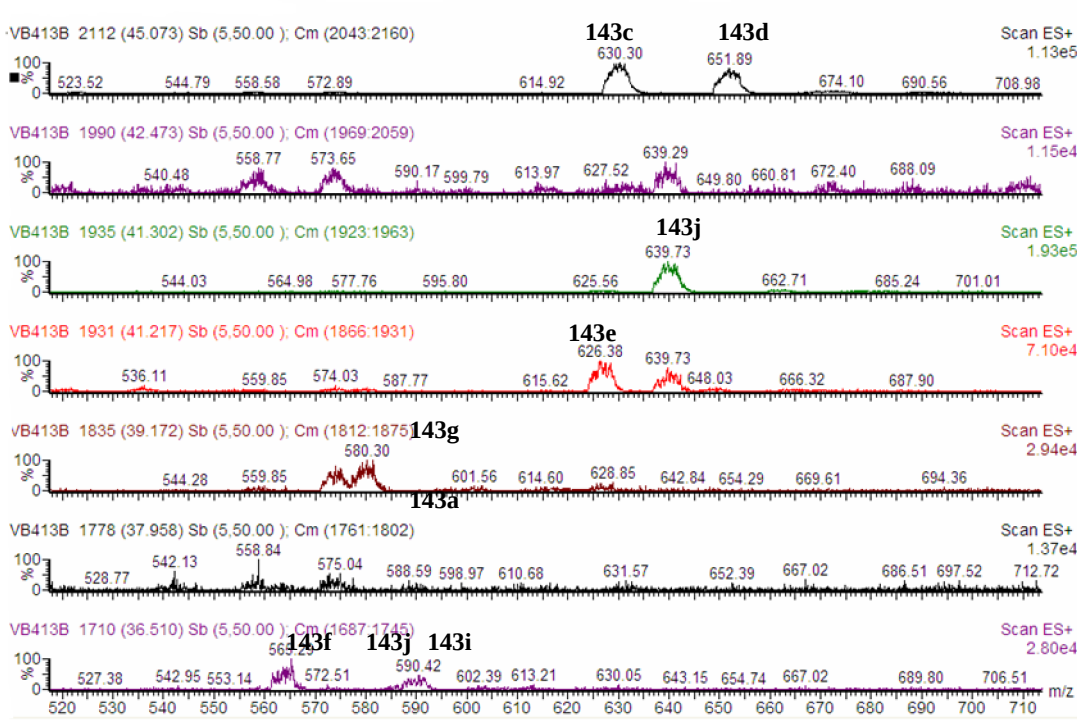


Figure 2.20 *MS data for DCL based on aldehyde* ***138****, indicating that all the expected ten acylhydrazones are present*

2.6.13 Isothermal Titration Calorimetry (experiments carried out by Anne Caniard)

Isothermal titration calorimetry (ITC) was carried out in collaboration with Pr. Alan Cooper and the Glasgow Biological Microcalorimetry Facility. The ITC measurements were performed on a VP-ITC calorimeter (Microcal Inc Northampton,

USA) at 25 °C. SjGST and hGSTP1-1 were dialysed against a 0.1 M potassium phosphate buffer, pH 6.8. The final dialysis buffer was used to make up the ligand solution as well as for instrument calibration and baseline controls. Hydrazones **143c** and **143g** were dissolved and titrated in 10 % DMSO. The concentrations of SjGST and hGSTP1-1 (approximately 20 μM) were determined by measuring the absorption at 280 nm and the concentrations of ligands were approximately 30 times higher.

The reactants SjGST and hGSTP1-1 were place in the 2 mL sample chamber and **143c** or **143g** in the syringe. A typical ITC measurement consisted of a first control injection of 1 μl followed by 29 successive injections of 10 μL for 20 s with a 3 min interval between each injection.

Raw data were collected, corrected for ligand heats dilution, and the peaks generated integrated using ORIGIN software (Microcal Inc) by plotting the values in microcalories against the molar ratio of injectant to reactant within the cell. Data were fitted using the one single-site binding model. From the dissociation constant K_D and the reaction enthalpy value ΔH, the change in free Gibbs energy (ΔG°) and entropy change (ΔS°) can be calculated using the equation $\Delta G° = -RT \ln(1/K_D) = \Delta H - T\Delta S°$ where R is the universal gas constant and T the absolute temperature.

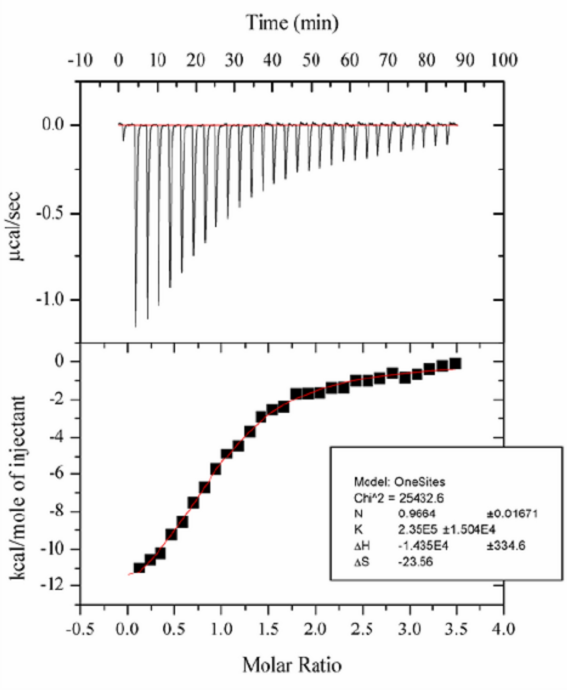


Figure 2.21a *ITC of SjGST in the presence and in the absence of* ***143g****. Top: Heats of injection of 276 µM* ***143g*** *into a cell containing 17 µM SjGST. Bottom: Data from the upper panel integrated and plotted as a function of the molar ratio of* ***143g*** *after subtraction of heats generated by injection of 276 µM* ***143g*** *into buffer*

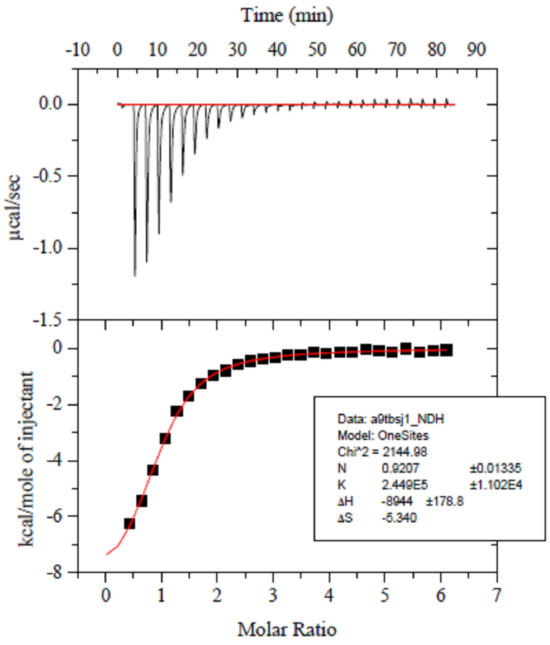


Figure 2.21b *ITC of SjGST in the presence and in the absence of* ***143c****. Top: Heats of injection of 604 µM* ***143c*** *into a cell containing 21 µM SjGST. Bottom: Data from the upper panel integrated and plotted as a function of the molar ratio of* ***143c*** *after subtraction of heats generated by injection of 604 µM* ***143c*** *into buffer.*

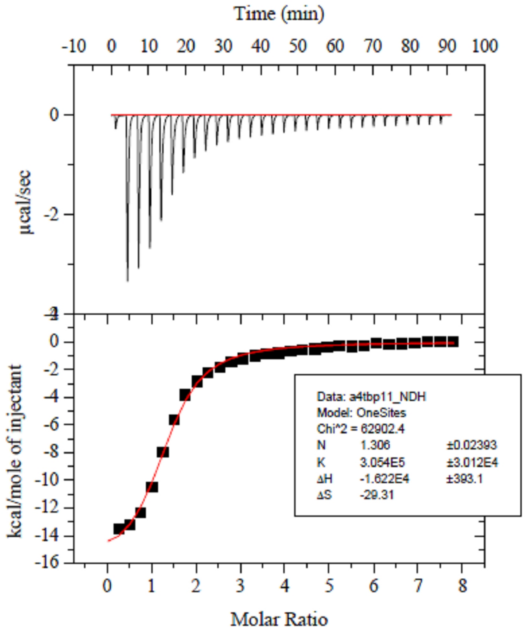


Figure 2.21c *ITC of hGSTP1-1 in the presence and in the absence of* ***143c****. Top: Heats of injection of 680 µM* ***143c*** *into a cell containing 20 µM hGSTP1-1. Bottom: Data from the upper panel integrated and plotted as a function of the molar ratio of* ***143c*** *after subtraction of heats generated by injection of 680 µM* ***143c*** *into buffer.*

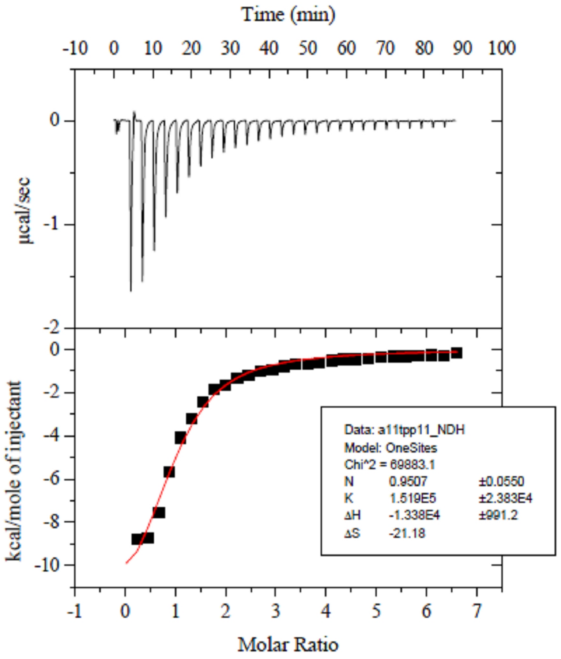
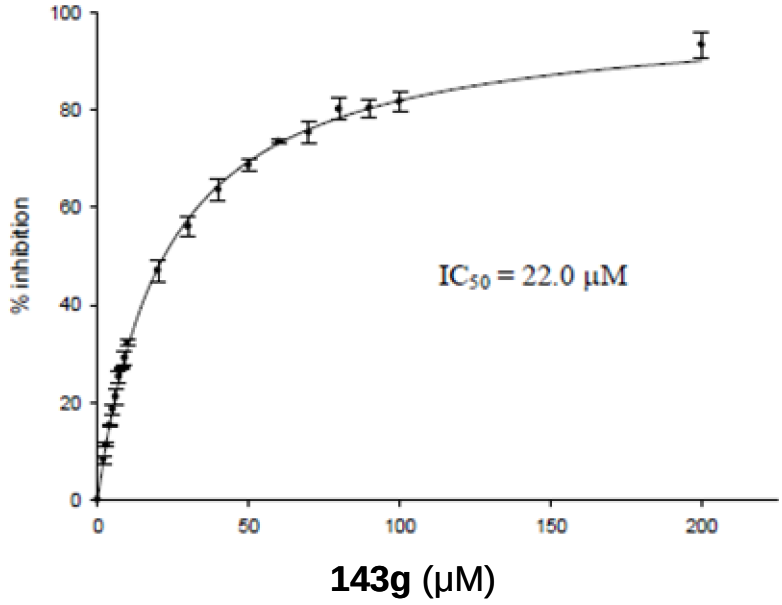


Figure 2.21d *ITC of hGSTP1-1 in the presence and in the absence of* ***143g****. Top: Heats of injection of 598 µM* ***143g*** *into a cell containing 20 µM hGSTP1-1. Bottom: Data from the upper panel integrated and plotted as a function of the molar ratio of* ***143g*** *after subtraction of heats generated by injection of 598 µM* ***143g*** *into buffer.*

2.6.14 Inhibition Assays (Experiments carried out jointly with Anne Caniard)

Inhibition assays were carried out on a Perkin Elmer Wallac VICTOR2™ 1420 Multilabel counter UV-visible 96 well-plate spectrophotometer. To 360 µl well were added phosphate buffer (255 µl, 0.1 M, pH 6.8), GST (15 µl, 0.15 $mg.ml^{-1}$) and inhibitor (15 µl, 0.2 to 200 µM). The solution was mixed well and after incubation at 25 °C for 5 min, CDNB (15 µl, 40 mM) and GSH (15 µl, 40 mM) were added quickly and mixed. Absorbance was measured at 340 nm, 25 °C for 5 min.

The K_i values with respect to both GSH and CDNB were determined. Inhibition kinetics were studied at 8 different, fixed concentrations of either GSH or CDNB (8, 16, 31, 63, 125, 250, 500 and 1000 µM) with variable concentrations of the inhibitor (2 to 200 µM). Data were collected and analysed using SigmaPlot® software.



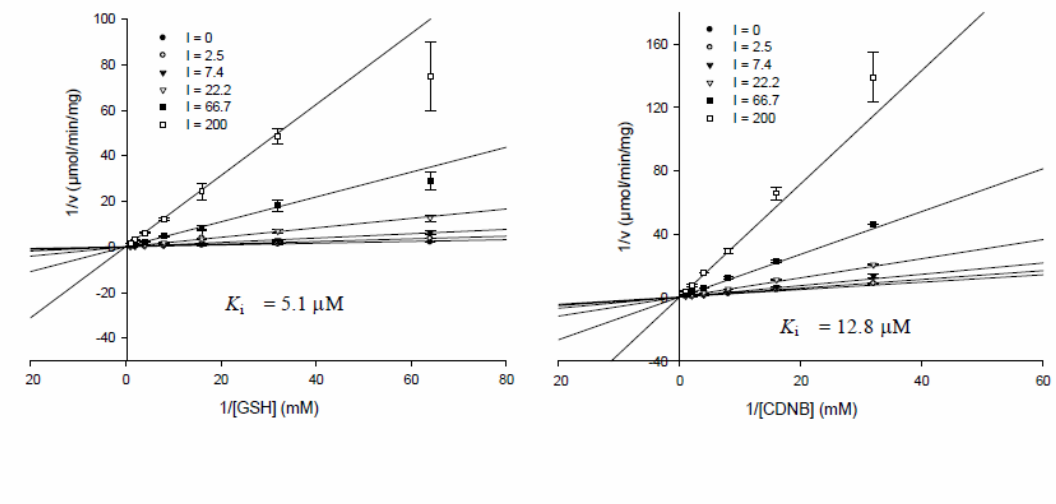


Figure 2.22a *Inhibition of SjGST by* ***143g*** *using GSH and CDNB as substrates. The IC_{50} value is the concentration of inhibitor giving 50% inhibition of enzyme activity. Data are the mean ± standard deviation of triplicate experiments.*

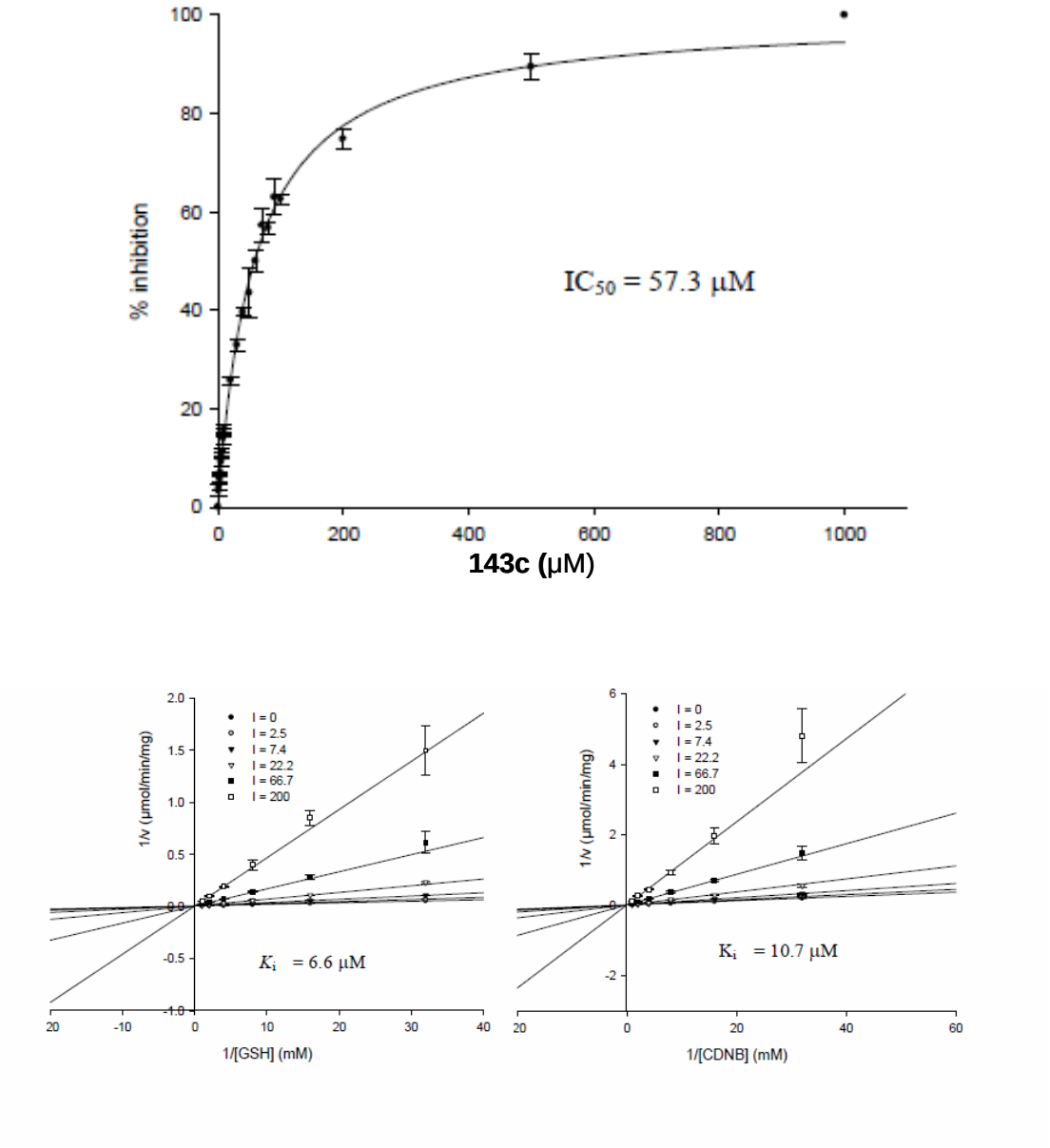


Figure 2.22b *Inhibition of hGSTP1-1 by* ***143c*** *using GSH and CDNB as substrates.*

2.6.15 **Protein Synthesis** (Anne Caniard and Dominic Campopiano)

Expression and purification of SjGST

The SjGST construct pGEX-6P-1 (GE Healthcare) was used to transform *E. coli* expression strain BL21 (DE3) and grown on LB/ampicillin (100 µg/ml) agar plates. One colony was used to inoculate 200 ml of LB broth containing ampicillin (100 µg/ml) and grown overnight at 37 °C with agitation (250 rpm). The 200 ml overnight culture was used to inoculate 3 L of LB broth and grown to an OD_{600} of 0.6 at 37 °C and then induced with 0.1 mM IPTG for 3 to 4 hours at 37 °C. Cells were harvested by refrigerated centrifugation at 5 000 x *g* for 10 min immediately after induction. Cell pellets were frozen and stored at -20 °C until required.

The SjGST expressing pellet was resuspended (4 ml/g wet cell pellet) in buffer binding buffer (50 mM Tris-HCl, pH 7) containing one protease inhibitor cocktail tablet (EDTA-free). The resuspended pellet was sonicated for 15 min (30 s on / 30 s off) on ice and was then centrifuged at 27 000 x *g* for 30 min at 4 °C to remove insoluble debris. The supernatant was filtered through a 0.45 µm membrane prior to chromatography at 4 °C. The cell lysate was loaded onto a 20 ml GSTPrep FF 16/10 column (GE Healthcare) previously equilibrated with binding buffer. The column was then washed with 5 column volumes of binding buffer before elution using a linear gradient of 0-100 % elution buffer (50 mM Tris-HCl, 10 mM glutathione, pH 8) over 5 column volumes. Fractions were analysed by SDS-PAGE and those containing sjGST were pooled and applied onto a 320 ml HiPrep Sephacryl S-200 HR column (GE Healthcare) pre-equilibrated with a 0.1 M potassium phosphate buffer containing 150 mM NaCl, pH 6.8. The column was then eluted with one column volume of the same buffer. Fractions containing sjGST were pooled and dialysed against 4 L of a 0.1 M potassium phosphate buffer, pH 6.8 at 4 °C.

Expression and purification of hGSTP1-1

Overexpression of His_6-Tagged hGSTP1-1 was carried out by transforming *E. coli* BL21 (DE3) cells with the pET15b/hGSTP1 vector. A single colony was added to 200 ml LB supplemented with ampicillin (100 $\mu g.ml^{-1}$) and grown overnight at 37 °C and 250 rpm. This seed culture was then used to inoculate 3 L of fresh LB medium and grown to OD_{600} of 0.6 before induction with isopropyl thio-β-$_D$-glalactoside

(IPTG, 1.0 mM final concentration) for 3 to 4 hours at 37 °C and 250 rpm. The cells were then harvested by centrifugation (5000 x *g* for 15 min at 4 °C) and stored at -20 °C. Cells overexpressing His_6-hGSTP1-1 were resuspended in binding buffer (50 mM Tris- HCl, 0.5 M NaCl, 5 mM imidazole, pH 7, 4 ml per gram of wet cell paste) with one protease inhibitor cocktail tablet (EDTA-free) and disrupted by sonication (15 pulses of 30 s at 30 s intervals) at 4 °C. The cell debris were removed by centrifugation at 27 000 x *g* for 30 minutes at 4 °C, after which the supernatant was filtered through a 0.45 µm membrane prior to chromatography. The cell lysate was loaded onto a 5 ml HisTrap™ HP column (GE Healthcare) previously equilibrated with the same binding buffer. The column was then washed with 5 column volumes of binding buffer before the bound material was eluted using a linear gradient of 0-100 % elution buffer (50 mM Tris- HCl, 0.5 M NaCl, 0.5M imidazole, pH 8) over 20 column volumes at 4 °C. Fractions were analysed by SDS-PAGE and those containing hGSTP1-1 were pooled and applied onto a 320 ml HiPrep Sephacryl S-200 HR column (GE Healthcare) pre-equilibrated with a 0.1 M potassium phosphate buffer containing 150 mM NaCl, pH 6.8. The column was then eluted with one column volume of the same buffer. Fractions containing hGSTP1-1 were pooled and dialysed against 4 L of a 0.1 M potassium phosphate buffer, pH 6.8 at 4 °C.

Site-directed mutagenesis, expression and purification of SjGST-Y7F

Plasmid pGEX-6P-1 was used as a template in the site-directed mutagenesis experiment. The Y7F-For (5’ -TCCCCTATACTAGGTTTTTGGAAAATTAAG- 3’) and Y7F-Rev (5’ -CTTAATTTTCCAAAAACCTAGTATAGGGGA- 3’) oligonucleotides were used to change the TAT codon in the SjGST sequence to TTT, so that the tyrosine amino-acid residue was replaced by a phenylalanine (SjGST-Y7F). Mutations were performed by using Stratagene site-directed mutagnenesis kit. Each reaction contained plasmid DNA (5 µl), 10x reaction buffer (5 µl), primer-forward (1 µl), primer-reverse (1 µl), dNTP mix (1 µl), distilled water (final volume 50 µl) and Pfu DNA polymerase 2.5 u / µl (1 µl). The amplification parameters were as follow: an initial DNA denaturation at 95 °C for 2 min followed by 16 cycles of denaturation at 95 °C for 30 sec, annealing at 55 °C for 60 sec and extension at 68 °C for 7 min. One microliter (5 U) of *Dpn*I restriction enzyme was added to the PCR

product and incubated at 37 °C for 2 h. The ligation product was used to transform competent *E. coli* cells DH5α (DE3). The new mutated plasmid was confirmed by DNA sequencing and named pGEX-6-P1-Y7F.

The pGEX-6P-1-Y7F construct was used to transform *E. coli* expression strain BL21 (DE3) and grown on LB/ampicillin (100 µg/ml) agar plates. One colony was used to inoculate 200 ml of LB broth containing ampicillin (100 µg/ml) and grown overnight at 37 °C with agitation (250 rpm). The 200 ml overnight culture was used to inoculate 3 L of 2 x YT broth and grown to an OD_{600} of 0.6 at 37 °C and then induced with 0.1 mM IPTG for 5 to 6 hours at 30 °C. Cells were harvested by refrigerated centrifugation at 5 000 x *g* for 10 min immediately after induction. Cell pellets were frozen and stored at -20 °C until required.

The Y7F SjGST expressing pellet was resuspended (4 ml/g wet cell pellet) in buffer binding buffer (50 mM Tris-HCl, pH 7) containing one protease inhibitor cocktail tablet (EDTA-free). The resuspended pellet was sonicated for 15 min (30 s on / 30 s off) on ice and was then centrifuged at 27 000 x *g* for 30 min at 4 °C to remove insoluble debris. The supernatant was filtered through a 0.45 µm membrane prior to chromatography at 4 °C. The cell lysate was loaded onto a 20 ml GSTPrep FF 16/10 column (GE Healthcare) previously equilibrated with binding buffer. The column was then washed with 5 column volumes of binding buffer before elution using a linear gradient of 0-100 % elution buffer (50 mM Tris-HCl, 10 mM glutathione, pH 8) over 5 column volumes. Fractions were analysed by SDS-PAGE and those containing sjGST were pooled and applied onto a 320 ml HiPrep Sephacryl S-200 HR column (GE Healthcare) pre-equilibrated with a 0.1 M potassium phosphate buffer containing 150 mM NaCl, pH 6.8. The column was then eluted with one column volume of the same buffer. Fractions containing sjGST were pooled and dialysed against 4 L of a 0.1 M potassium phosphate buffer, pH 6.8 at 4 °C.

2.6.16 Protein analyses (Anne Canniard and Dominic Campopiano)

Mass spectrometry analyses of the pure enzymes gave the molecular masses of the *Schistosoma japonicum* GST as 28 434 Da, the His_6- hGSTP1-1 as 25 629 Da and the SjGST-Y7F as 28 410 Da in good agreement with the predicted values of 28 428 Da, 25 630 Da and 28 412 Da for the monomers, respectively. The concentration of SjGST, hGSTP1-1 and SjGST-Y7F were determined using a Bradford assay. The proteins were stored in a 0.1 M potassium phosphate buffer, pH 6.8 containing 20 % glycerol (v/v) at -20 °C.

2.6.17 Molecular modelling (carried out by Dr. Ruth Brenk and Dr. Torsten Luksch at Dundee University).

Ligand alignment: The superposition of glutathione S-transferase ligands was carried out by using Relibase+3.0.0.[45] First, a search was performed to find binding sites that share a sequence identity between 40 % and 100 % to the target GST crystal structure 1m9a. The resulting 38 structures with bound ligand were superimposed by using binding site residues only. Finally, the ligands from the superimposed structures were extracted and visually analysed.

Binding mode prediction with Moloc[46]

The SjGST crystal structure (pdb code: 1m9a SjGST – S-hexyl-GSH complex) and the hGST P1-1 – GSH complex crystal structure (pdb code: 6gss) were used as starting conformations for binding mode generation. The glutathione groups of the synthesized ligands were mapped onto the glutathione groups of the ligands bound to the crystal structures. The hydrophobic hydrazone groups of the synthesized ligands were oriented towards the cavity, lying at the end of the S-hexyl site, as observed for the EPNP ligand bound to cGSTM1-1 (pdb code 1c72). In the next step, the protein in complex with the modeled ligand was minimized, considering the ligand as fully flexible. For the protein all residues were kept rigid, except the amino acids that define the pocket at the end of the S-hexyl site (R103, V106, V161, V162, Q204 for SjGST and R100, Y103, I161, H162, N204 for hGST P1-1).

2.7 Notes and References

1 Ramström, O.; Lehn, J.-M. *ChemBioChem* **2000**, *1*, 41-48.

2 Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *J. Am. Chem. Soc.* **2000**, *122*, 12063-12064.

3 Shi, B.; Greaney, M. F. *Chem. Commun.* **2005**, 886-888.

4 Shi, B.; Stevenson, R.; Campopiano, D. J.; Greaney, M. F. *J. Am. Chem. Soc.* **2006**, *128*, 8459-8467.

5 Huc, I.; Lehn, J.-M. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 2106-2110.

6 Cousins, G. R. L.; Poulsen, S.-A.; Sanders, J. K. M. *Chem. Commun.* **1999**, 1575-1576.

7 Bunyapaiboonsri, T.; Ramström, O.; Lohmann, S.; Lehn, J.-M.; Peng, L.; Goeldner, M. *ChemBioChem* **2001**, *2*, 438-444.

8 Bunyapaiboonsri, T.; Ramström, H.; Ramström, O.; Haiech, J.; Lehn, J.-M. *J. Med. Chem.* **2003**, *46*, 5803-5811.

9 Kalia, J.; Raines, R.T. *Angew, Chem. Int. Ed.* **2008**, *47*, 7523-7526.

10 Furlan R. L. E.; Ng, Y.-F.; Otto S.; Sanders, J. K. M. *J. Am. Chem. Soc.* **2001**, *123*, 8876–8877.

11 Roberts, S. L.; Furlan, R. L. E.; Cousins G. R. L.; Sanders, J. K. M. *Chem. Commun.* **2002**, 938-939.

12 Liu, J.; West, K. R.; Bondy, C. R.; Sanders J. K. M. *Org. Biomol. Chem.* **2007**, *5*, 778-786.

13 Poulsen, S.A. *J. Am. Soc.Mass Spectrom.* **2006**, *17*, 1074-1080.

14 Skene, W.G.; Lehn, J.-M. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 8270-8275.

15 Cordes, E. H.; Jencks, W. P. *J. Am. Chem. Soc.* **1962**, *84*, 826-831.

16 Dirksen, A.; Dirksen, S.; Hackeng, T. M.; Dawson, P. E. *J. Am. Chem. Soc.* **2006**, *128*, 15602–15603.

17 Dirksen, A.; Dawson, P. E. *Bioconjugate Chem.* **2008**, *19*, 2543-2548.

18 Dirksen, A.; Hackeng, T. M.; Dawson, P. E. *Angew, Chem. Int. Ed.* **2006**, *45*, 7581- 7584.

19 Hayes, J. D.; Flanagan, J. U.; Jowsey, I. R. *Annu. Rev. Pharmacol. Toxicol.* **2005**, *45*, 51-88.

20 Mahajan, S.; Atkins, W. M. *Cell. Mol. Life Sci.* **2005**, *62*, 1221-1233.

21 Armstrong, R.N. *Curr. Opin. Chem. Biol*. **1998**, *2*, 618-623.

22 Mannervik, B.; Danielson, U.H. *CRC Crit. Rev. Biochem.* **1988**, *23*, 283-337.

23 Dirr, H.; Reinemer, P.; Huber, R. *Eur. J.Biochem.* **1994**. *220*, 645-661.

24 Dourado, D.F.; Fernandes, P.A.; Mannervik, B.; Ramos, M. J. *Chemistry* **2008**, *14*, 9591-9598.

25 Ketley, J.N.; Habig, W.H.; Jakoby, W.B. *J. Biol. Chem*, **1975**,*250*, 8670-8673.

26 Tew, K. D. *Cancer Res.* **1994**, *50*, 6449-6454.

27 Jao, S.-C.; Chen, J.; Yang, K.; Li, W.-S. *Bioorg. Med. Chem.* **2006** *14*, 304-318.

28 Lyon, R. P.; Hill, J. J.; Atkins, W. M. *Biochemistry* **2003**, *42*, 10418-10428.

29 Armstrong, R.N. *Chem. Re.s Toxicol*. **1997**, *10*, 2-18.

30 Manoharan, H.T.; Gulick, A.M.; Puchalski, R.B.; Servais, A.L.; Fahl, W.E. *J.Biol.Chem.* **1992**, *267*, 18940-18945.

31 Stenberg, G.; Board, P.G.; Mannervik, B.. *FEBS Lett*. **1991,** *293***,** 153-155.

32 Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J-L.; Sanders, J. K. M.; Otto, S. *Chem. Rev.* **2006**,*106*, 3652-3711.

33 Dynamic Combinatorial Chemistry in Drug Discovery, Bioorganic Chemistry, and Materials Science; edited by Miller, B.L.; Wiley, 2010.

34 Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Curr. Opin. Chem. Biol.* **2002**, *6*, 321-327.

35 Cheeseman, J.D.; Corbett, A. D.; Cleason, J.L.; Kazlauskas, R.J. *Chem. Eur. J.* **2006**, *11*, 1708-1716.

36 Rowan, S.J.; Cantrill, S.J.; Cousins, G.R.L.; Sanders, J.K.M.; Stoddart, J.F. *Angew, Chem. Int. Ed.* **2002**, *41*, 898-906.

37 Zhao, G.; Yu,T.; Wang,R.; Wang, X.; Jing,Y. *Bioorg. Med. Chem.* **2005**, *13*, 4056-62.

38 Oakley, A.J.; Rossjohn, J.; Lo Bello, M.; Caccuri, a.M.; Fredrici, G.; Parker, M.W. *Biochemistry* **1997**, *36*, 576-585.

39 Oakley, A.J.; Lo Bello, M.; Mazetti, A.P.; Fredrici,G.; Parker, M.W. *FEBS Lett.* **1997**, *419*, 32-36.

40 Andújar-Sánchez, M.; Smith, A.W.; Clemente-Jimenez, J. M.; Rodriguez-Vico, F.; Las Heras-Vazquez, F. J.; Jara-Pérez, V.; Cámara-Artigas, A. *Biochemistry* **2005,** *44*, 1174-1183.

41 Campoy, A. Z.; Freire, E. *Nature Protocols* **2006**, *1*,186-191.

42 Ladbury, J.E.; Klebe, G.; Freire, E. *Nature Reviews:Drug Discovery* **2010**, *9*, 23-27.

43 Pierce, M.M.; Raman, C.S.; Nall, B.T. *Methods* **1999**, *19*, 213-221.

44 Cardoso, R. M.; Daniels, D. S.; Bruns, C. M.; Tainer, J. A. *Proteins* **2003**, *51*, 137-146.

45 Bergner, A.; Günther, J.; Hendlich, M.; Klebe, G.; and Verdonk, M. *Biopolymers* **2001**, *61*, 99-110.

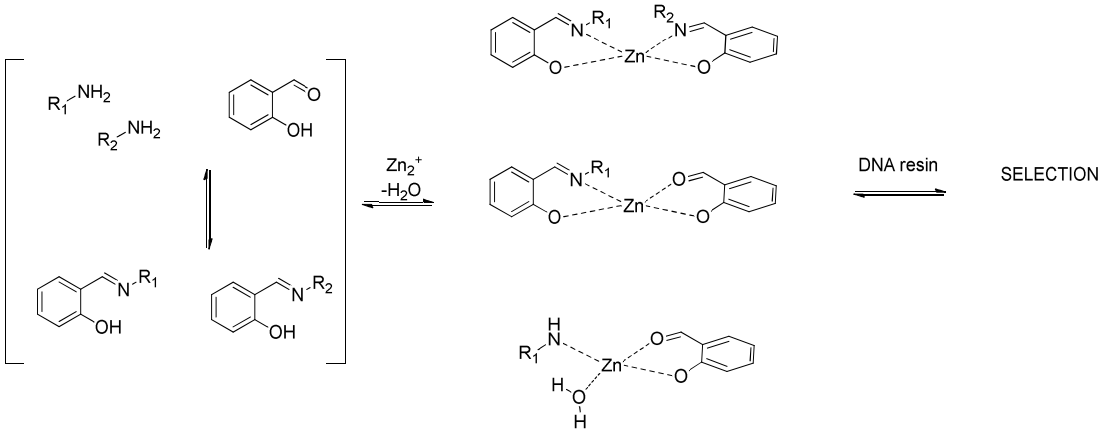
46 Gerber, P. R.; Müller, K. *J. Comput. Aided Mol. Des.* **1995**, *3*, 251-268.

Chapter 3

Multi-level dynamic systems

3.1 Introduction

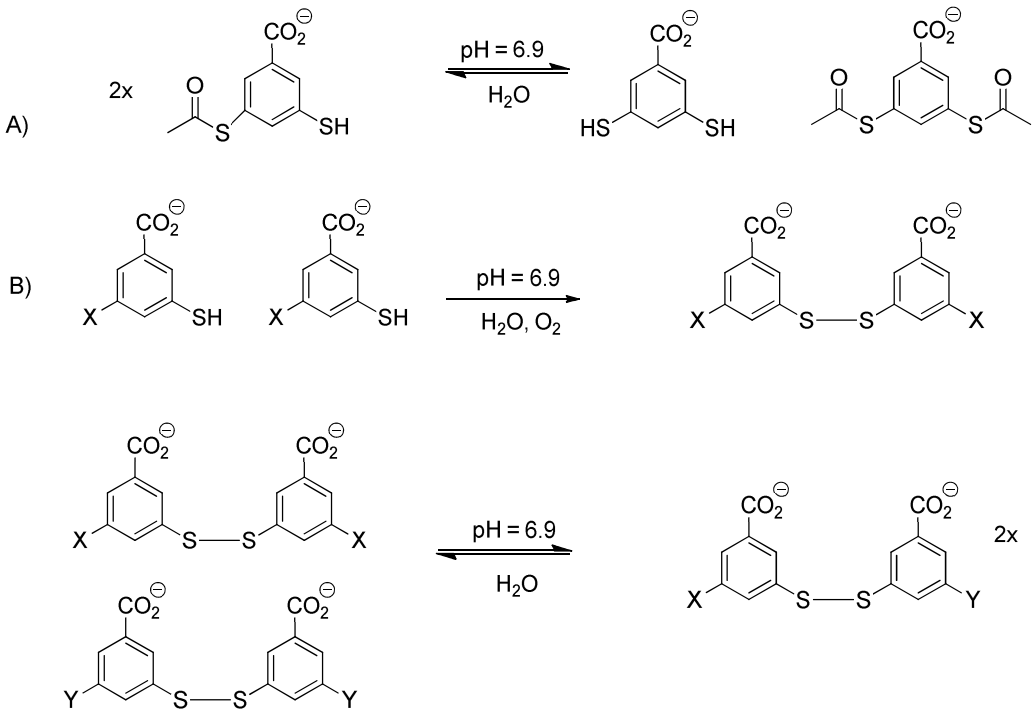
Recently, dynamic combinatorial libraries (DCLs)[1,2] have become an invaluable tool in the study of emergent behaviour of complex interacting molecular networks, an area now referred to as systems chemistry.[3-10] The choice of the reversible reaction is a fundamental design parameter in constructing dynamic combinatorial libraries: the exchange process has to be rapid under mild conditions and should not interfere with molecular recognition events which aid library evolution. Reactions commonly used include disulfide, hydrazone/imine and thioester exchange. More complex systems which combine two or more reversible reactions to generate multilevel dynamic libraries with increased diversity have appeared recently in literature.[11-18] One of the earliest examples using multiple equilibria was reported by Miller and Klekota in 1999 (Scheme 3.1). They used imine formation and transition- metal complexation between Zn(II) ions and salicylaldimines to generate DCLs targeting DNA.[11]



Scheme 3.1 *DCLs using imine formation and metal complexation targeting DNA*

Sanders, Otto and co-workers have reported a double- level 'communicating' library based on the simultaneous exchange of disulfide and thioester bonds in aqueous solution at neutral pH (Scheme 3.2).[13] A complex system involving three exchange processes incorporating disulfides, imines and metal coordination has also been

reported by Nitschke.[15] Furlan *et al* have combined disulfide, hydrazone and thioester chemistry in a single system. [19]

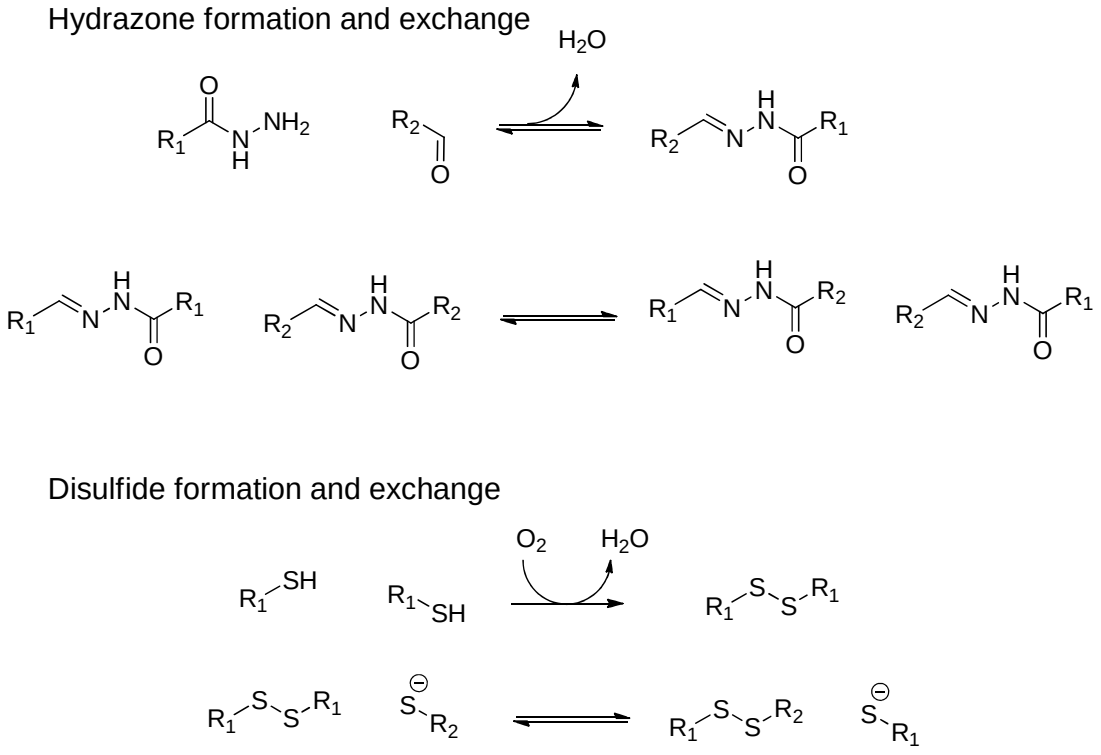


Scheme 3.2 *A double- level 'communicating' DCL based on the simultaneous exchange of disulfide and thioester bonds*

Orthogonal double-level libraries, in which one exchange process is selectively activated while the other is kinetically slow under the conditions employed, are an interesting variation. Hydrazone exchange coupled orthogonally to ligand exchange around a cobalt ion has been reported by Eliseev and Lehn.[12] In this system ligand exchange at the terpyridine cobalt(II) centre can be turned off by oxidation of the metal to cobalt(III) while imine exchange can be turned off by increasing the pH. The combinations of imine and nitrone exchange [20] and hydrazone exchange with azobenzene photoisomerisation [21] have also been reported.

A recent addition to the list of orthogonal dynamic processes is the combination of disulfide and hydrazone exchange in a single system reported simultaneously by Otto

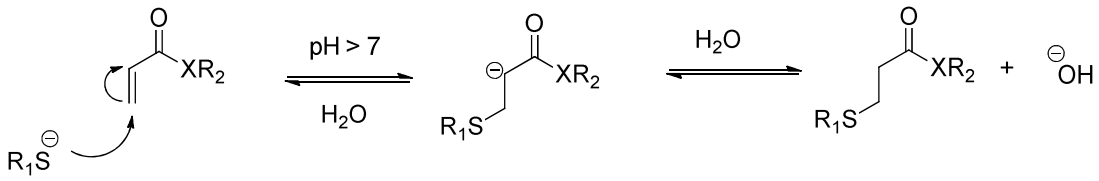
and Furlan (Scheme 3.3).[17,18] The hydrazone exchange process is exclusively activated under acidic conditions while basic conditions switch on the disulfide exchange. The unique features of this orthogonal system has been used by Leigh *et al* to construct a synthetic small molecule walker that can walk down a track.[22,23] This fully synthetic model of the cellular motor protein kinesin emphasizes the advantages of orthogonal dynamic covalent chemistry and led us to focus on the discovery of new orthogonal reaction systems.



Scheme 3.3 *Reactions involved in hydrazone and disulfide exchange*

We have explored here the combination of hydrazone exchange[24-30] and Michael addition of thiols to enones as two independently addressable reversible reactions in a single aqueous dynamic system. Previous work in the Greaney group had established reversible Michael addition of thiols to enones under mild aqueous conditions as a tool for DCC (Scheme 3.4).[31] The reaction was successfully applied later to generate DCLs targeting the enzyme glutathione-S-transferase (GST) in an

attempt to identify best-binding ligands.[32] The reaction equilibrates rapidly under basic conditions but is extremely slow at low pH. On the other hand, fast hydrazone exchange requires acid catalysis (pH < 4) and the reaction is irreversible under basic conditions.[33,34] The aim of this project was to establish the optimal experimental conditions required to facilitate orthogonal use of these reversible reactions in aqueous solution.



Scheme 3.4 *Reversible Michael addition of thiols to enones*

3.2 Results and Discussion

In order to prove the orthogonality of the two reactions, we began by preparing a DCL of mono-functional components that would set up independent acyl hydrazone and conjugate addition exchange processes (Figure 3.1). We chose the enone ethacrynic acid (EA) **73** as our Michael acceptor, previously used by us for targeting the enzyme GST. EA derivatives equilibrate quickly (< 3 hr) with the cellular thiol glutathione (GSH) **69** at pH 8.[31,32]

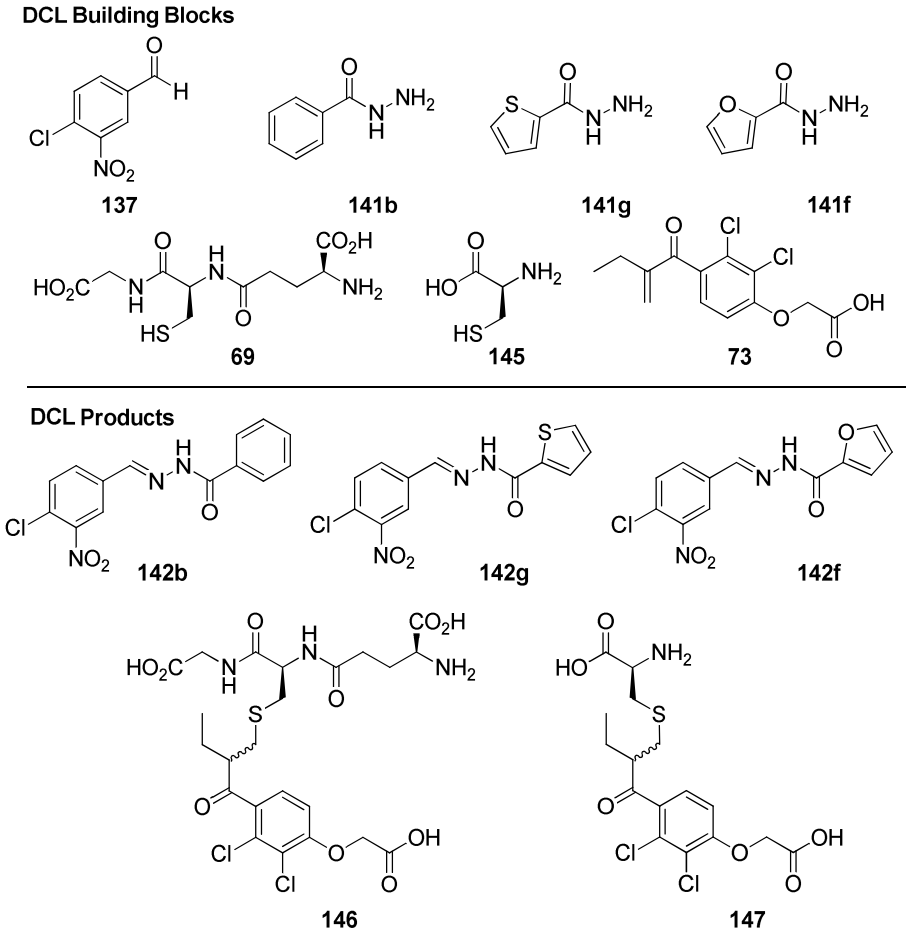


Figure 3.1 *DCL components: the orthogonal DCL is built from one aldehyde, three hydrazides, two thiols and one enone.*

3.2.1 pH dependence of conjugate addition of thiols to enones

For the orthogonal DCL to be successful, we require thiol conjugate addition to out-compete any intefering conjugate addition of the nucelophilic hydrazides.[35] To probe this chemoselectivity, we mixed enone **73** with two samples containing a mixture of GSH (**69**), and hydrazide **141b** under acidic and basic conditions. At pH 8.5 Michael adduct **146** formed immediately even though an equivalent of hydrazide was present in the system. Under acidic conditions (pH 3.5), however, no reaction occurred and the concentration of **73** remained constant over 6 hours. An experiment with L-cysteine (**145**) instead of thiol **69** gave similar results (Figure 3.2).

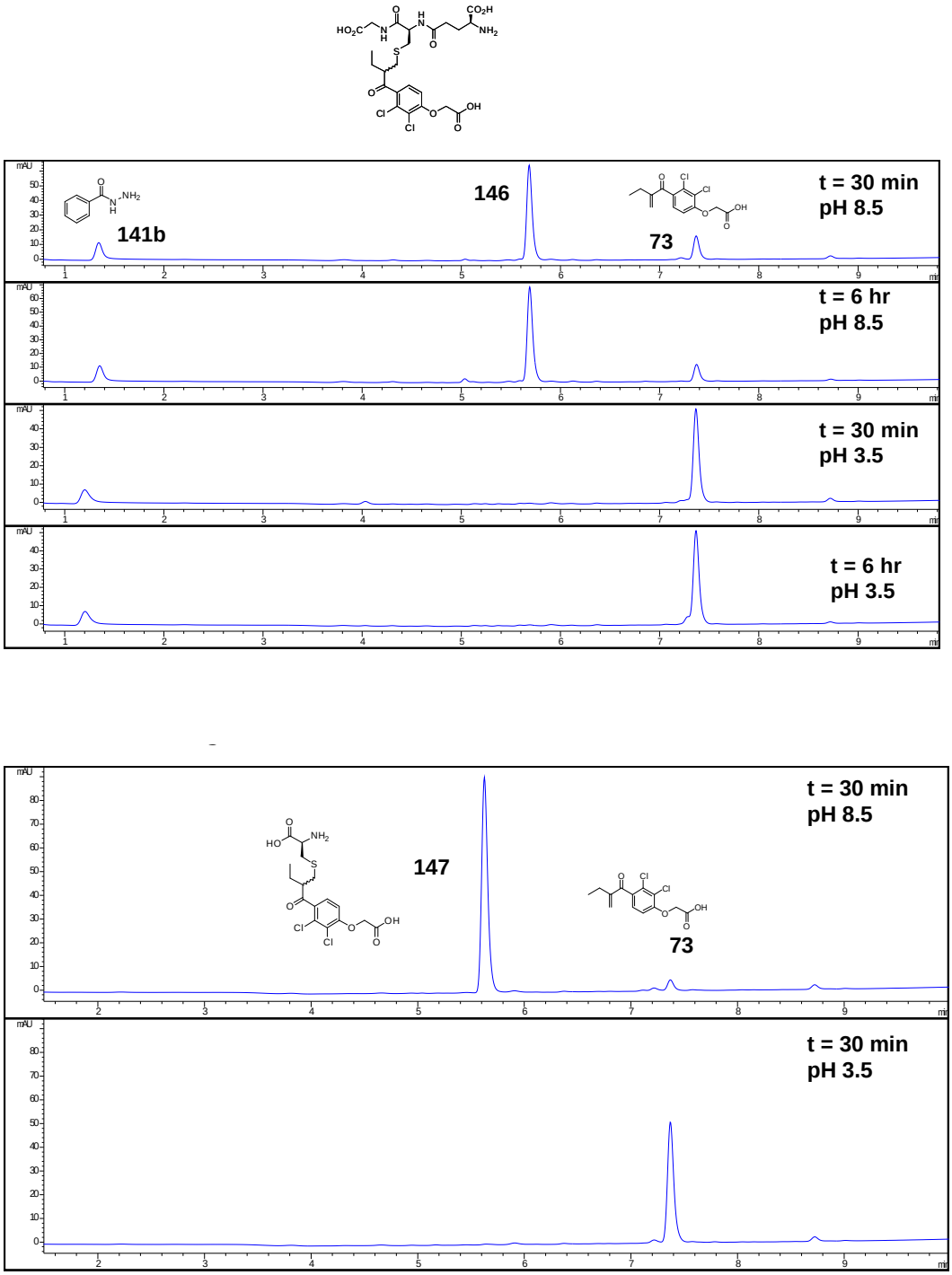


Figure 3.2 *pH dependence of conjugate addition of thiols to enones Ethacrynic acid (**73**) (5 µL, 10 mM, DMSO), glutathione (**69**) (10 µL, 10 mM, aqueous) and hydrazide **141b** (10 µL, 10 mM, DMSO) were added to a mixture of DMSO (185 µL) and ammonium acetate buffer (790 µL, 50 mM, pH 8.5) Another sample was similarly set up in ammonium acetate buffer pH 3.5 and monitored over 6 hours. Ethacrynic acid (**73**) (5 µL, 10 mM, DMSO) and L-cysteine (**145**) (10 µL, 10 mM, aqueous) were added to a mixture of DMSO (195 µL) and ammonium acetate buffer (790 µL, 50 mM, pH 8.5). All samples were analysed by HPLC at 254 nm.*

3.2.2 pH dependence of hydrazone formation

The reverse chemoselectivity, that of hydrazides out-competing thiols in reaction with aldehydes, was confirmed by the same protocol where Michael acceptor **73** is replaced by aldehyde **137**. Rapid hydrazone formation was observed at pH 3.5 even in the presence of the thiol. No hydrazone was formed in 30 minutes at pH 8.5 though small amounts of **142b** were detected when the system was allowed to react over 6 hours (Figure 3.3)

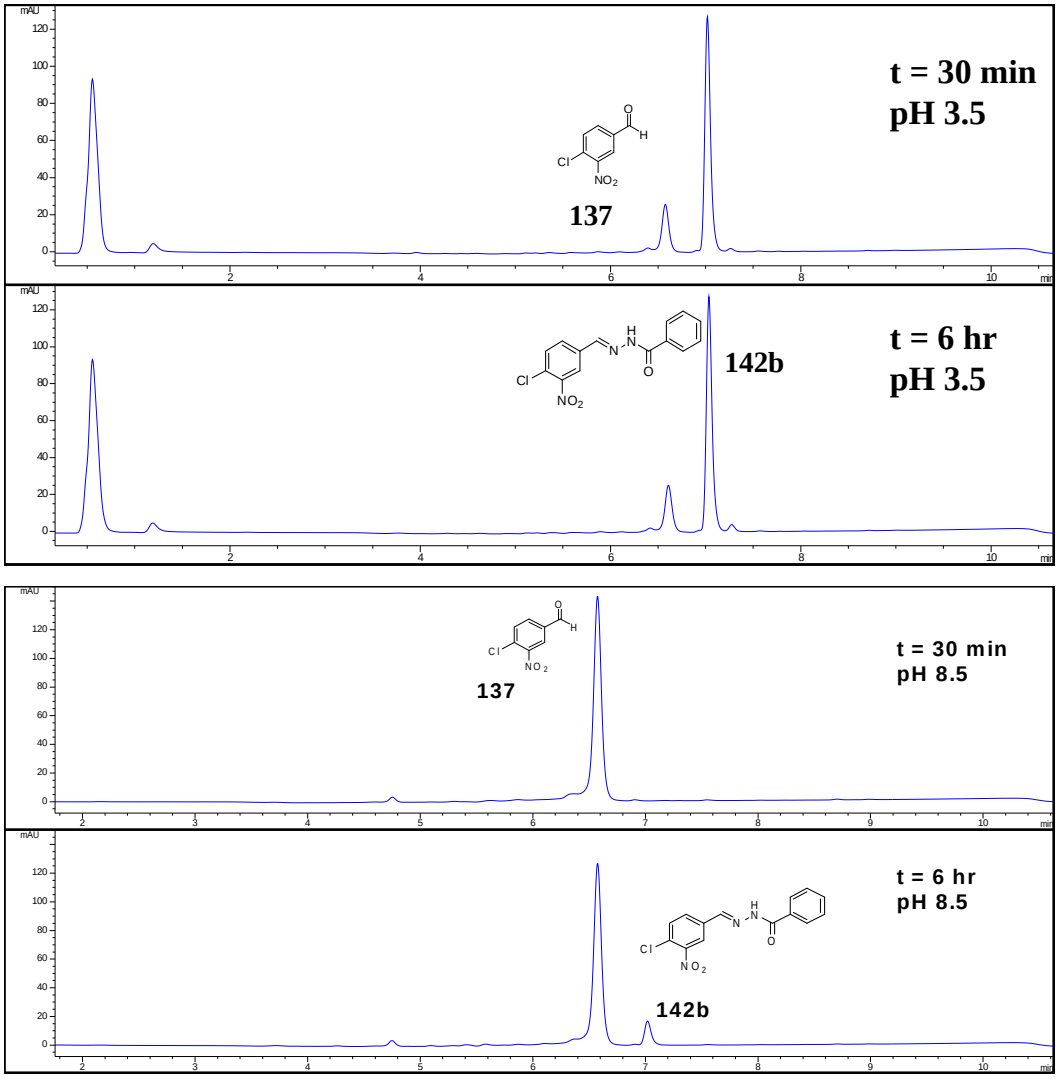


Figure 3.3 *pH dependence of hydrazone formation. Aldehyde **137** (5 µL, 10 mM, DMSO), glutathione (**69**) (10 µL, 10 mM, aqueous) and hydrazide **141b** (10 µL, 10 mM, DMSO) were added to a mixture of DMSO (185 µL) and ammonium acetate buffer (790 µL, 50 mM, pH 3.5). A similar sample was mixed at pH 8.5. All samples analysed by HPLC at 254 nm.*

3.2.3 Monofunctional conjugate addition-hydrazone exchange

To adapt these results to a pH-addressable DCL that could be switched between the two channels of reactivity, we mixed together aldehyde **137** (50 µM), enone **73** (50 µM), GSH **69** (100 µM) and hydrazides **141b** and **141g** (100 µM each) in 20% DMSO-ammonium acetate buffer (50 mM, pH 8.5) at room temperature. Analysing the mixture by HPLC at 254 nm after 30 min showed the formation of the EA-GSH conjugate **146** as expected while no detectable amounts of hydrazones **142b** and **142g** were found. (Figure 3.4, trace A1) We then deactivated the conjugate addition by switching the pH to 3.5 with acetic acid. An immediate sampling of the mixture by HPLC indicated appreciable formation of the hydrazones **142b** and **142g** and a reduction in the aldehyde **137** (Figure 3.4,trace A2). Subsequent analysis over time showed that the electron-deficient aldehyde **137** was completely consumed in less than 6 hours with the hydrazones **142b** and **142g** in a dynamic equilibrium (Figure 3.4, trace A3). Importantly, no change in concentrations of the Michael acceptor **73** or the adduct **146** was observed throughout this period indicating that the reaction was switched off.

We next decided to probe the reversibility of the hydrazone exchange under conditions where the Michael addition is deactivated (Figure 3.4, traces B1 – B5). We added excess of hydrazide **141f** (200 µM) to the sample mixture from experiment A at pH 3.5. Hydrazone **142f** started forming immediately though no free aldehyde was detected in the system.(Figure 3.4, traces B1-B3) Monitoring the mixture over 12 hours showed that the hydrazone equilibrium had shifted considerably towards **142f** with substantially lower amounts of **142b** and **142g** now present. Most gratifyingly, the concentrations of Michael acceptor **73** and adduct **146** remained constant throughout. Further, adding the more reactive thiol cysteine (**145**, 100 µM**)** to the mixture under the same acidic conditions did not change the concentration of **73** over 2 hours nor did the Michael adduct **147** appear, confirming the true orthogonality of the double-level dynamic system (Figure 3.4, trace B5).[36,37]

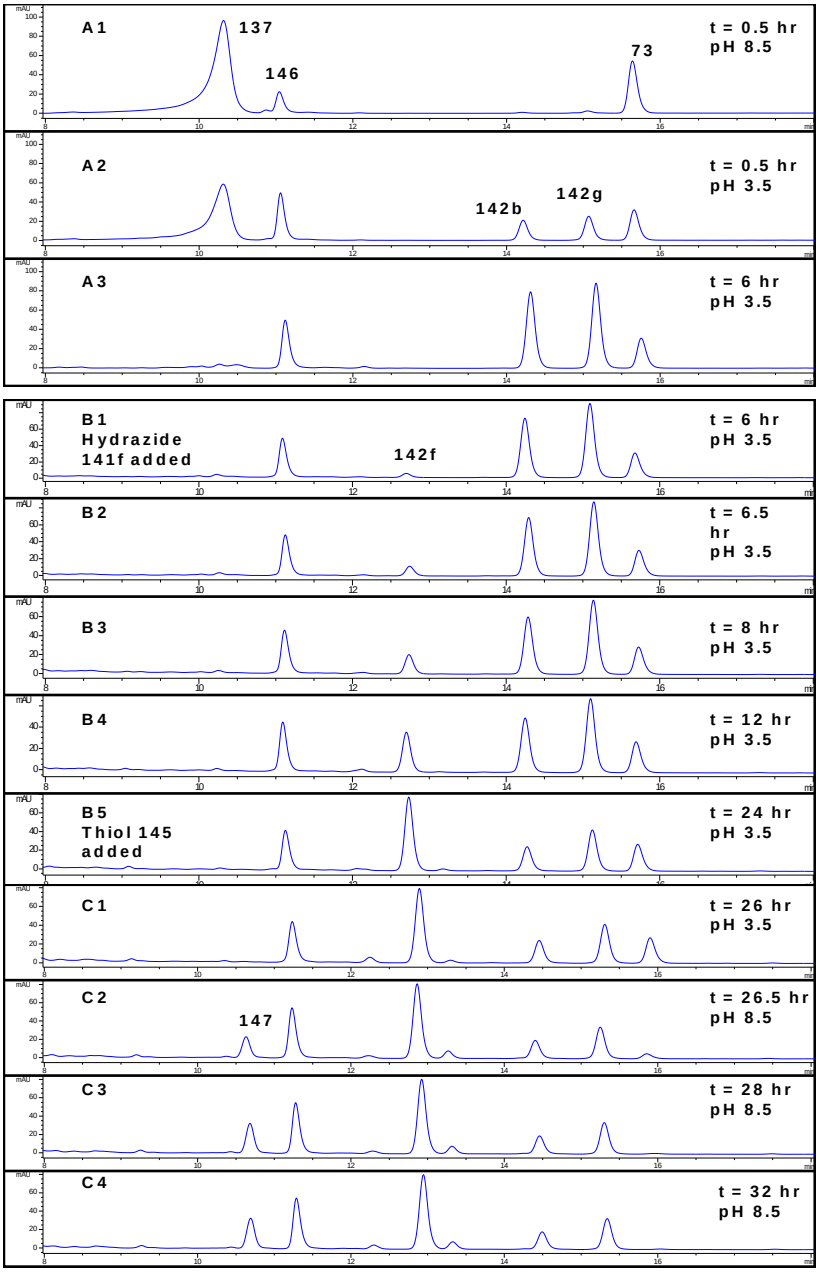


Figure 3.4 *pH-Addressable DCL. HPLC UV (254 nm) traces showing the elution profile of DCL product components derived from aldehyde* ***137*** *and enone* ***73****. The starting thiols and hydrazides are very polar and elute close to the solvent front in each analysis. The HPLC analysis is split into three cumulative phases: Phase A examines the pH orthogonality of hydrazone formation and conjugate addition. Phase B demonstrates adaptation of the hydrazone DCL in the presence of frozen conjugate addition components. Phase C demonstrates re-activation of conjugate addition chemistry following the switching off of hydrazone exchange.*

To probe the pH dependence of the hydrazone exchange we switched on the Michael addition by titrating the same mixture with 5M aqueous NaOH to pH 8.5. As expected, cysteine adduct **147** formed rapidly within 30 min with no EA **73** detected in the mixture after 2 hours (Figure 3.4, trace C2). Additionally, the concentrations of the hydrazones **142b**, **142g** and **142f** remained constant over 6 hours of reaction (Figure 3.4, traces C3-C4).

3.2.4 Bifunctional conjugate addition-hydrazone exchange

Orthogonal dynamic covalent chemistry using hetero-bifunctional molecules would be a significant advantage in generating complex molecular structures. To explore the utility of hydrazone formation and conjugate addition in this context we prepared a DCL based around the bifunctional thio-hydrazide **148** (Figure 3.5).

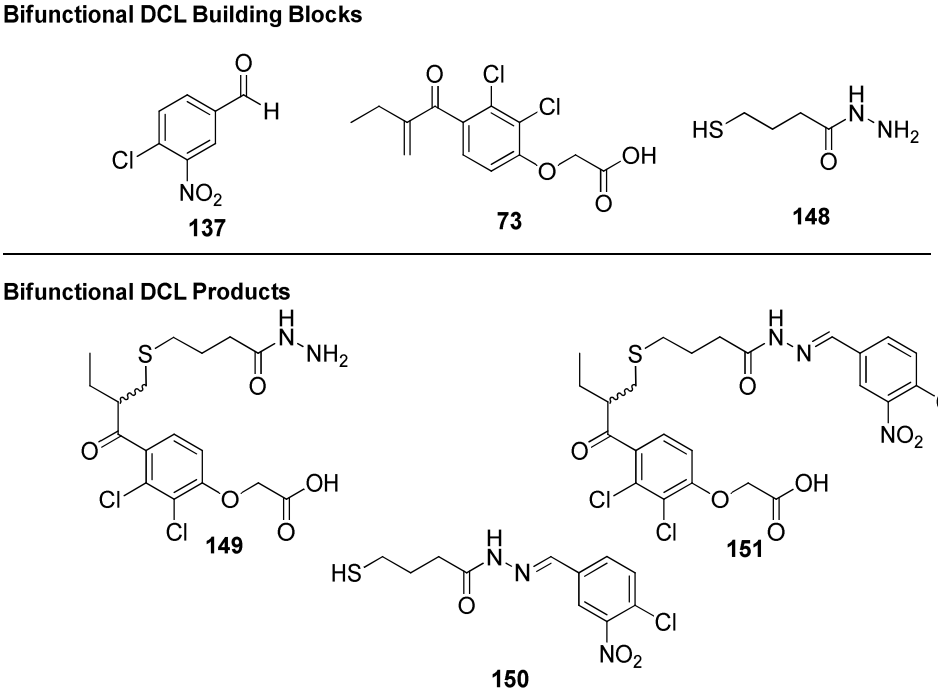


Figure 3.5 *Bifunctional DCL components*

Initially, compound **148** (100 µM) was added to a mixture of aldehyde **137** and enone **73** (50 µM each) in aqueous buffer containing 20% DMSO at pH 8.5. Monitoring the reaction over a period of 12 hours showed a progressive decrease in the concentration of **73** while the amount of aldehyde **137** remained constant throughout. The expected conjugate adduct **149** eluded detection under our LC conditions, although its presence in the mixture was confirmed by mass spectrometry (m/z = 437). Indirect evidence of its formation was obtained when we switched the pH of the mixture to 3.5. The concentration of aldehyde **137** rapidly reduced over time and significant amount of compound **151** was formed, indicating that the initial conjugate addition of the thiol in **148** had taken place (Figure 3.6).

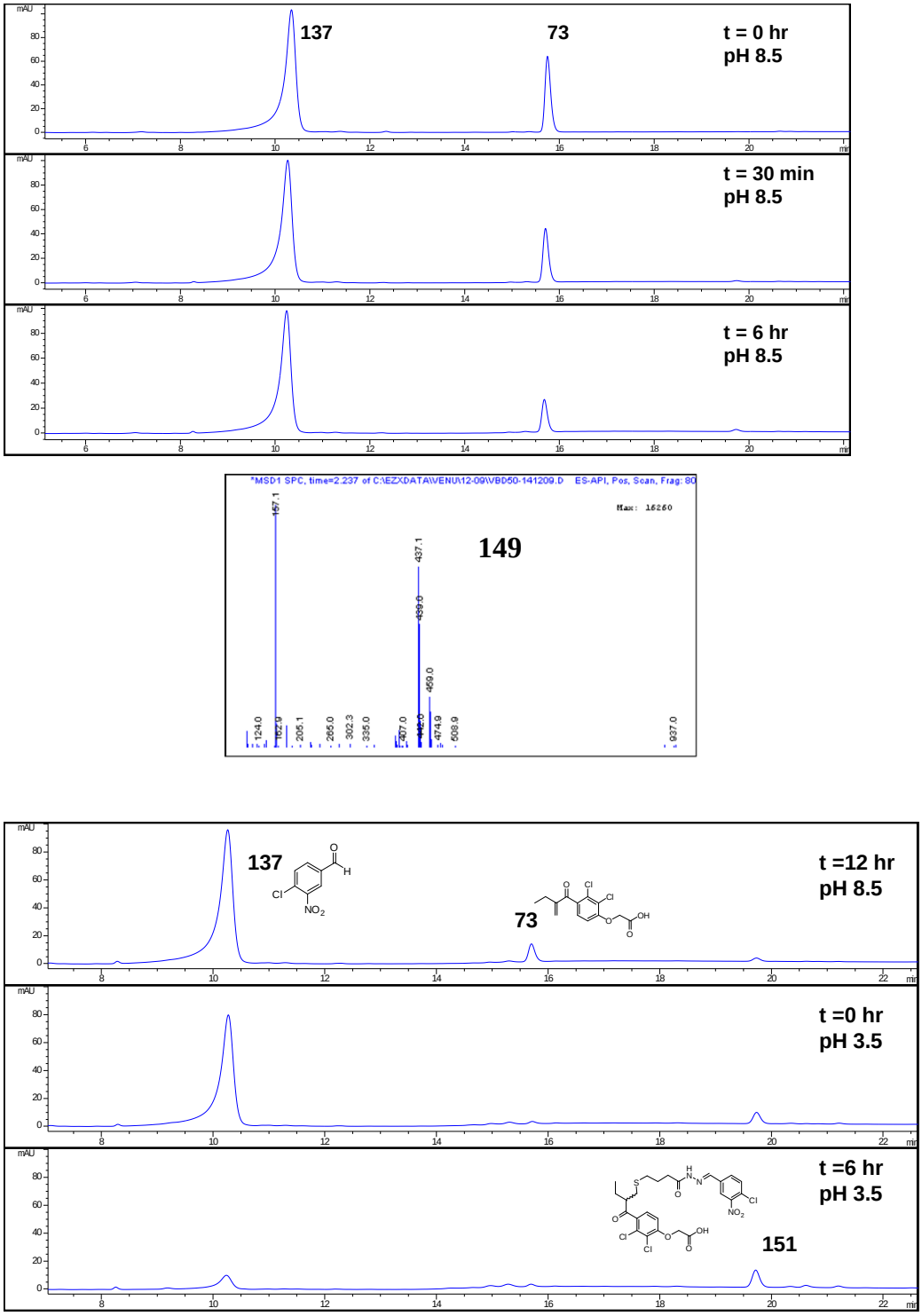


Figure 3.6 *Orthogonal conjugate addition and hydrazone formation with Compound* ***148.*** *Ethacrynic acid (**73**) (20 µL, 10 mM, DMSO), aldehyde **137** (20 µL, 10 mM, DMSO), and thiohydrazide **148** (40 µL, 10 mM, aqueous), were added to a mixture of DMSO (760 µL) and ammonium acetate buffer (3160 µL, 50mM, pH 8.5). Analysed byHPLC over 12 hr at 254 nM. The pH of the mixture was then switched to 3.5 by titration with glacial acetic acid and analysed over 6 hours by HPLC.*

To confirm pH control over a more complex system we mixed together aldehyde **137**, enone **73**, GSH **69**, thiohydrazide **148** and hydrazide **141b** (all 100 µM each) in DMSO-ammonium acetate buffer at pH 8.5 (Figure 3.7). Acidification and sampling after 5 min showed rapid formation of GSH adduct **146** had occurred, with adduct **149** not being detected as in the previous case. Analysis after 6 hours showed adduct **146** in stasis, the complete disapperance of aldehyde **137** and the formation of hydrazones **142b**, **150** and **161**.

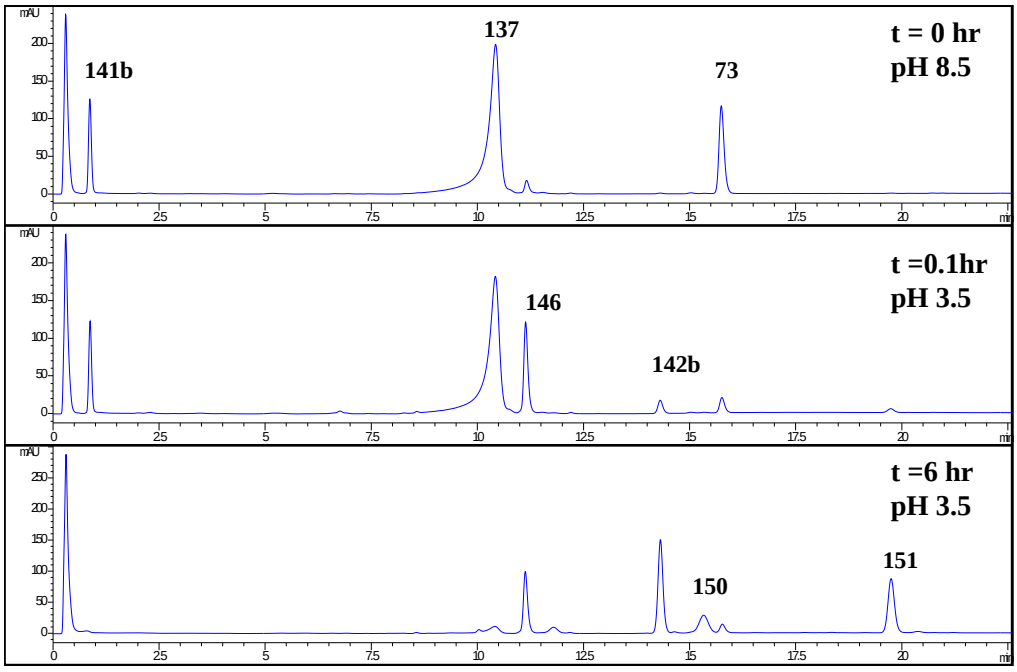


Figure 3.7 *Bifunctional DCL. Ethacrynic acid (**73**) (10 µL, 10 mM, DMSO), aldehyde **137** (10 µL, 10 mM, DMSO), glutathione (**69**) (10 µL, 10 mM, aqueous), thiohydrazide (**148**) (10 µL, 10 mM, DMSO) and hydrazide **141b** (10 µL, 10 mM, DMSO) were added to a mixture of DMSO (160 µL) and ammonium acetate buffer (790 µL, 50 mM, pH 8.5). Analysed by HPLC at 254 nM over 6 hours followed by a switch in pH of to 3.5 and then monitored by HPLC periodically over 6 hours.*

3.2.5 Disulfide formation under conjugate addition conditions

To confirm if disulfide formation competes with conjugate addition under our conditions we mixed together thiol **152** with ethacrynic acid **73** (100 µM each) at pH 8.5. Though thiol 152 formed the disulfide when it was on it own at pH 8.5, in the presence of the enone, Michael addition was extremely fast with almost no disulfide being detected by HPLC (Figure 3.8).

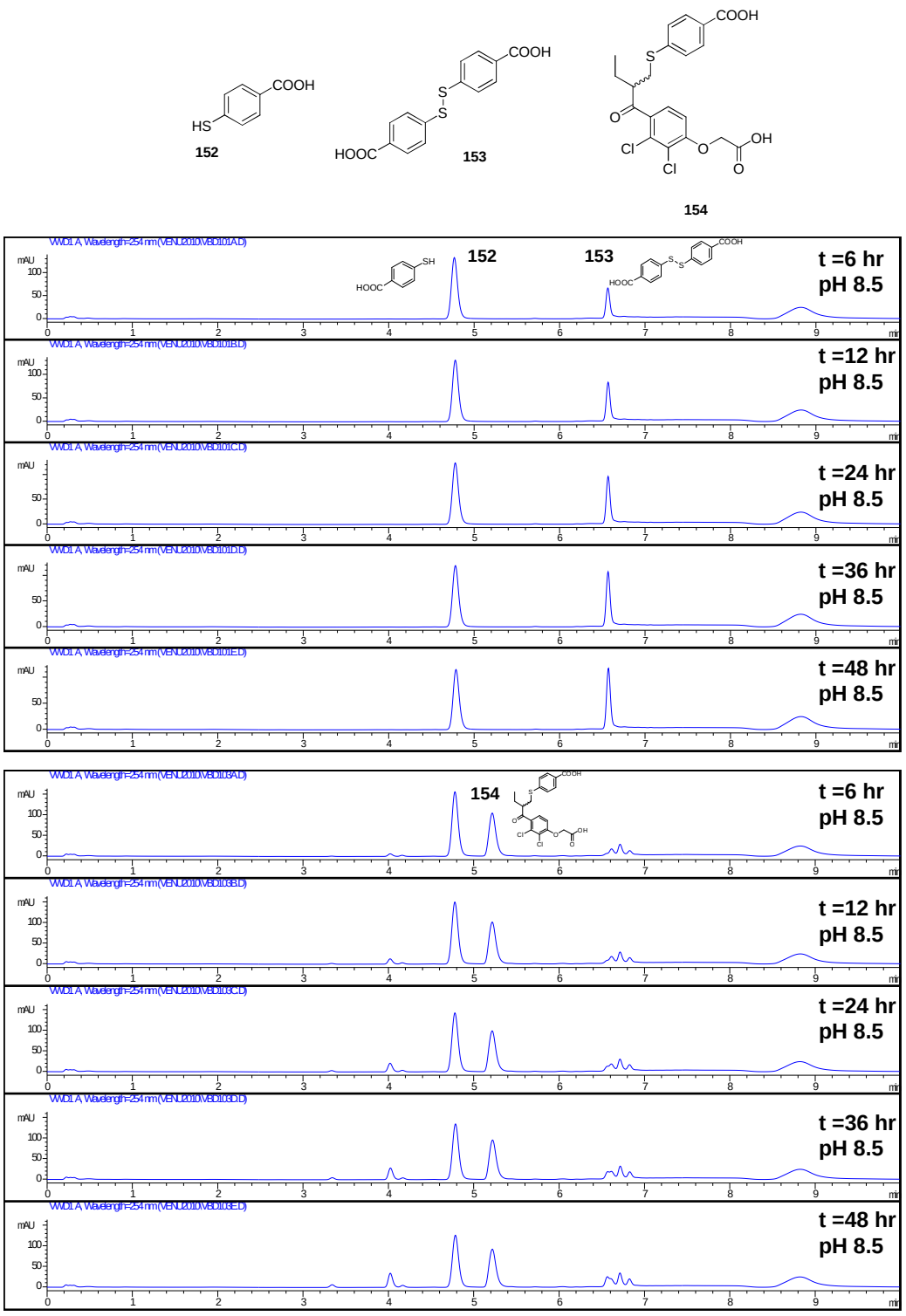


Figure 3.8 *Disulfide formation under conjugate addition conditions. 4-thiobenzoic acid (**152**) (10 µL, 10 mM, MeCN), was added to a mixture of MeCN (190 µL) and ammonium acetate buffer (800 µL, 50 mM, pH 8.5 and analysed by HPLC at 254 nM over 48 hours. Again, 4-thiobenzoic acid (**152**) (10 µL, 10 mM, MeCN), and ethacrynic acid (**73**) (5 µL, 10 mM, MeCN) and were added to a mixture of MeCN (185 µL) and ammonium acetate buffer (800 µL, 50 mM, pH 8.5 and analysed by HPLC at 254 nM over 48 hours.*

3.3 Conclusion

In conclusion, we have established that thiol conjugate addition and hydrazone formation exhibit very high levels of orthogonality when combined in a single DCC experiment. The DCLs are formed in aqueous solution and are easily controlled by simple pH change, undergoing cyclical de- and re-activation according to pH regime. From a design perspective, our work contrasts with recent orthogonal DCL systems in using exchange processes that feature heterofunctional reactivity, rather than homofunctional disulfide bond exchange.[17,18]

3.4 Experimental

3.4.1 General

All DCL components for the current study were purchased from commercial chemical companies with the exception of compound **148**, an initial sample of which was obtained from Max von Delius (Leigh group). HPLC analysis was carried out on a Agilent 1100 machine. Column: Luna 5 µ C18(2) 50 mm × 4.6 mm, flow rate : 1mL min^{-1}, wavelength 254 nm, temperature 30 ^{o}C, gradient H_2O / MeCN (0.01% TFA) from 95% to 5% over 30 min then back to 95% over 3 min.

3.4.2 pH Dependence of conjugate addition of thiols to enones

Ethacrynic acid (**73**) (5 µL, 10 mM, DMSO), glutathione (**69**) (10 µL, 10 mM, aqueous) and hydrazide **141b** (10 µL, 10 mM, DMSO) were added to a mixture of DMSO (185 µL) and ammonium acetate buffer (790 µL, 50 mM, pH 8.5) in 2 mL screw-cap vial. The sample composition was analysed by HPLC at 254 nM over 6 hours. Another sample was similarly set up in ammonium acetate buffer pH 3.5 and monitored over 6 hours.

Ethacrynic acid (**73**) (5 µL, 10 mM, DMSO) and L-cysteine (**145**) (10 µL, 10 mM, aqueous) were added to a mixture of DMSO (195 µL) and ammonium acetate buffer (790 µL, 50 mM, pH 8.5) in a 2 mL screw-cap vial. The sample composition was

analysed by HPLC at 254 nM after 30 min. Another sample was similarly set up in ammonium acetate buffer at pH 3.5 and monitored.

3.4.3 pH Dependence of hydrazone formation

Aldehyde **137** (5 µL, 10 mM, DMSO), glutathione (**69**) (10 µL, 10 mM, aqueous) and hydrazide **141b** (10 µL, 10 mM, DMSO) were added to a mixture of DMSO (185 µL) and ammonium acetate buffer (790 µL, 50 mM, pH 8.5) in a 2 mL screw-cap vial. The sample composition was analysed by HPLC at 254 nM over 6 hours. Another sample was similarly set up in ammonium acetate buffer pH 3.5 and monitored over 6 hours.

3.4.4 Monofunctional conjugate addition-hydrazone exchange

Experiment A

Ethacrynic acid (**73**) (20 µL, 10 mM, DMSO), aldehyde **137** (20 µL, 10 mM, DMSO), glutathione (**69**) (40 µL, 10 mM, aqueous) and hydrazides **141b** and **141g** (40 µL each, 10 mM, DMSO) were added to a mixture of DMSO (680 µL) and ammonium acetate buffer (3160 µL, 50 mM, pH 8.5) in a 7 mL screw-cap vial. The sample was analysed by HPLC at 254 nM after 30 min. The sample mixture was immediately titrated with glacial acetic acid to pH 3.5 and monitored by HPLC periodically over 6 hours.

Experiment B

An excess of hydrazide **141f** (80 µL, 10 mM, DMSO) was then added to the sample mixture from Experiment A (pH 3.5) and the composition analysed by HPLC over a period of 6 hours. L-cysteine (**145**) (40 µL, 10 mM, aqueous) was then added to the mixture at pH 3.5 and the composition was monitored periodically over 2 hours.

Experiment C

The pH of the sample mixture from Experiment B was now switched to 8.5 by titrating with aqueous 5M NaOH followed by periodic HPLC analysis over 6 hours.

3.4.5 Orthogonal conjugate addition and hydrazone formation with compound 148

Ethacrynic acid (**73**) (20 µL, 10 mM, DMSO), aldehyde **137** (20 µL, 10 mM, DMSO), and thiohydrazide **148** (40 µL, 10 mM, aqueous), were added to a mixture of DMSO (760 µL) and ammonium acetate buffer (3160 µL, 50mM , pH 8.5) in a 7 mL screw-cap vial. The sample was analysed by HPLC at 254 nM over 12 hours. The pH of the mixture was then switched to 3.5 by titration with glacial acetic acid and analysed over 6 hours by HPLC.

3.4.6 Bifunctional conjugate addition-hydrazone exchange

Ethacrynic acid (**73**) (10 µL, 10 mM, DMSO), aldehyde **137** (10 µL, 10 mM, DMSO), glutathione (**69**) (10 µL, 10 mM, aqueous), thiohydrazide (**148**) (10 µL, 10 mM, DMSO) and hydrazide **141b** (10 µL, 10 mM, DMSO) were added to a mixture of DMSO (160 µL) and ammonium acetate buffer (790 µL, 50 mM, pH 8.5) in a 2 mL screw-cap vial. The sample was analysed by HPLC at 254 nM over 6 hours. The pH of the sample mixture was then switched to 3.5 by titration with glacial acetic acid and monitored by HPLC periodically over 6 hours.

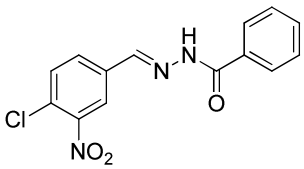
3.4.7 Disulfide formation under conjugate addition conditions

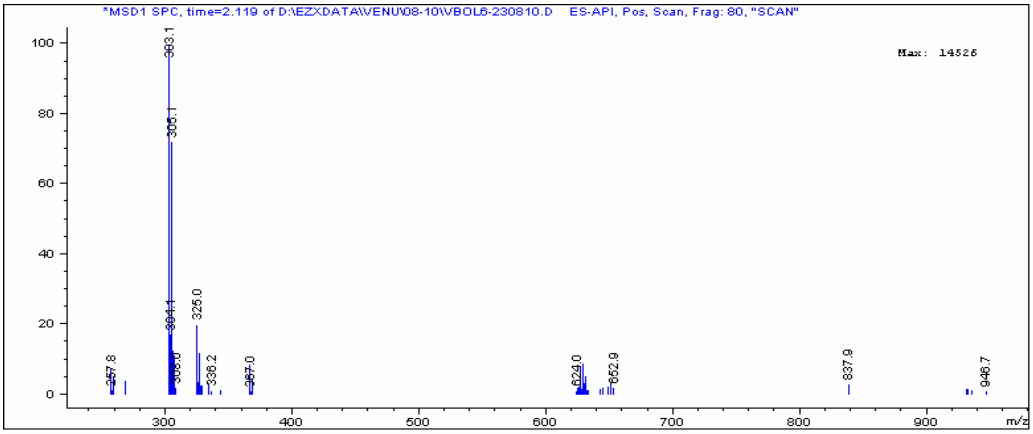
4-thiobenzoic acid (**152**) (10 µL, 10 mM, MeCN), was added to a mixture of MeCN (190 µL) and ammonium acetate buffer (800 µL, 50 mM, pH 8.5) in a closed 2 mL screw-cap vial. The sample composition was analysed by HPLC at 254 nM over 48 hours.

Again, 4-thiobenzoic acid (**152**) (10 µL, 10 mM, MeCN), and ethacrynic acid (**73**) (5 µL, 10 mM, MeCN) and were added to a mixture of MeCN (185 µL) and ammonium acetate buffer (800 µL, 50 mM, pH 8.5) in 2 mL screw-cap vial. The sample composition was analysed by HPLC at 254 nM over 48 hours.

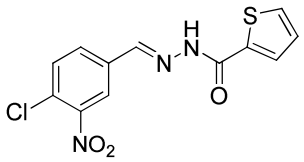
3.4.8 MS Data for the DCL compounds

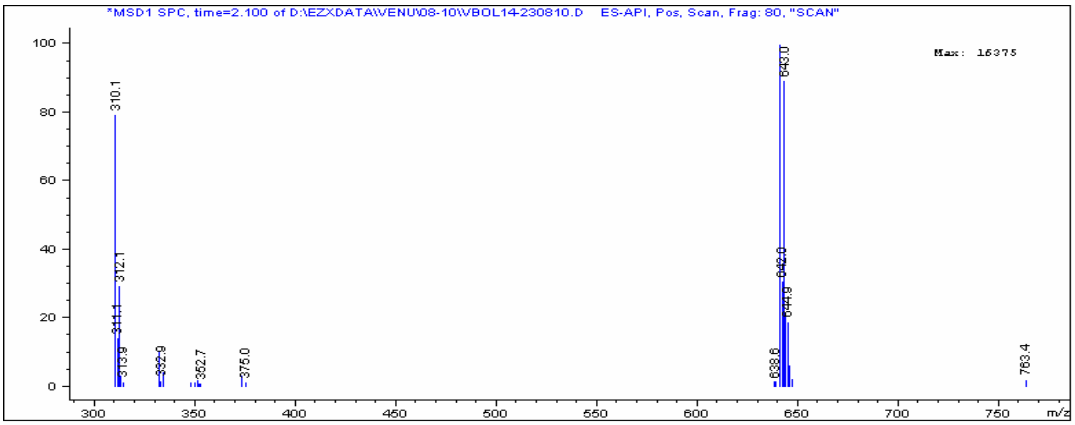
Compound 142b



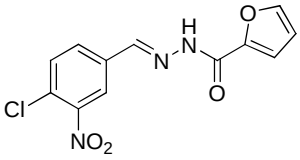


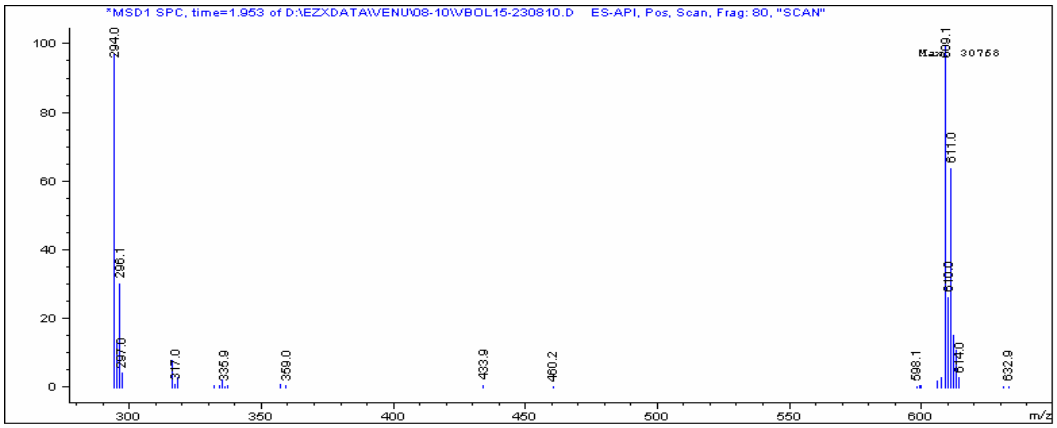
Compound 142g



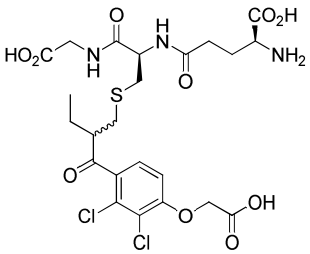


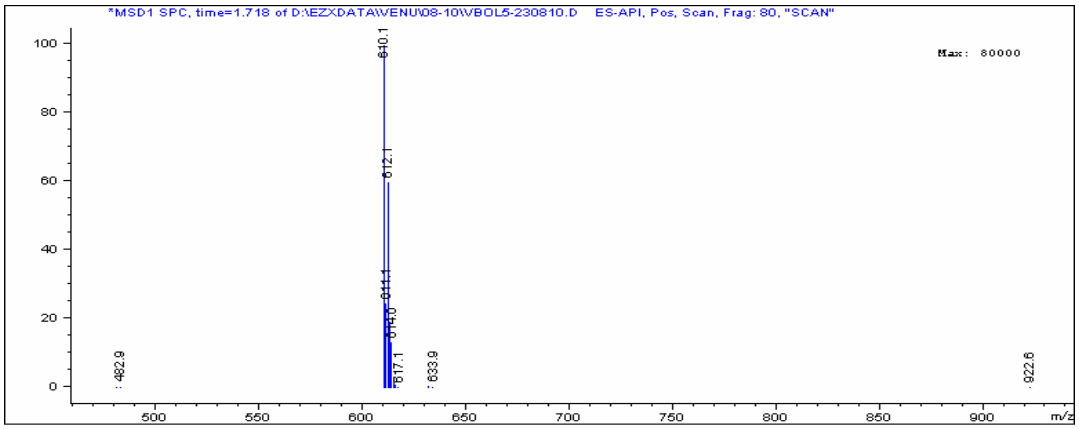
Compound 142f



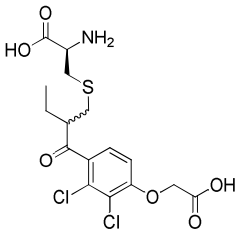


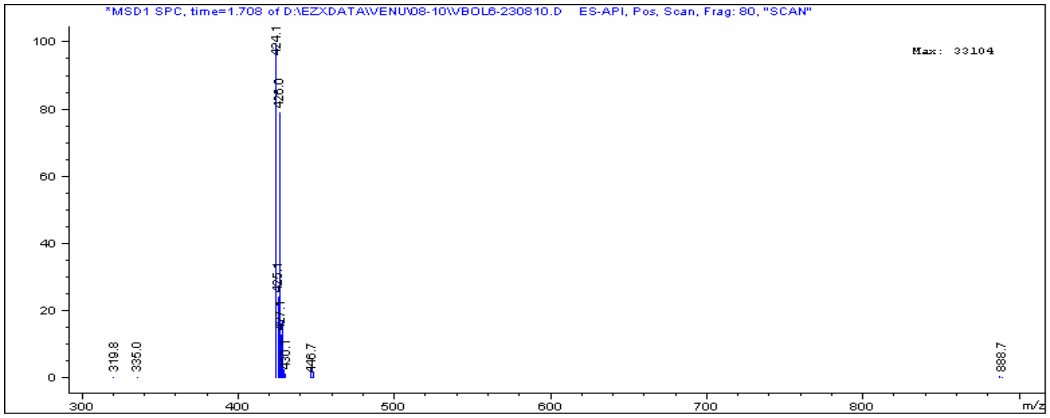
Compound 146



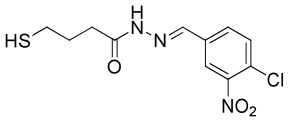


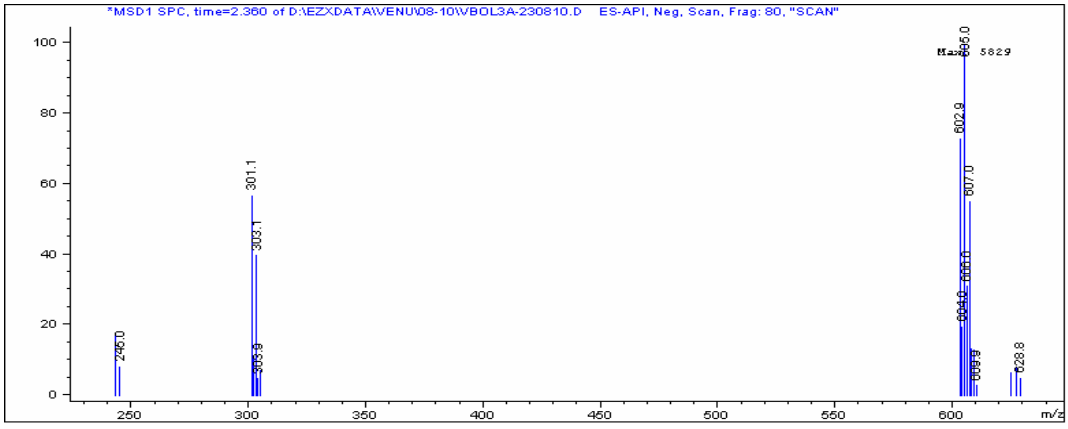
Compound 147



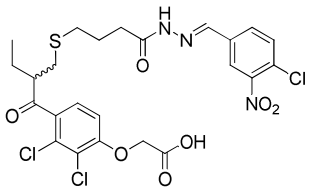


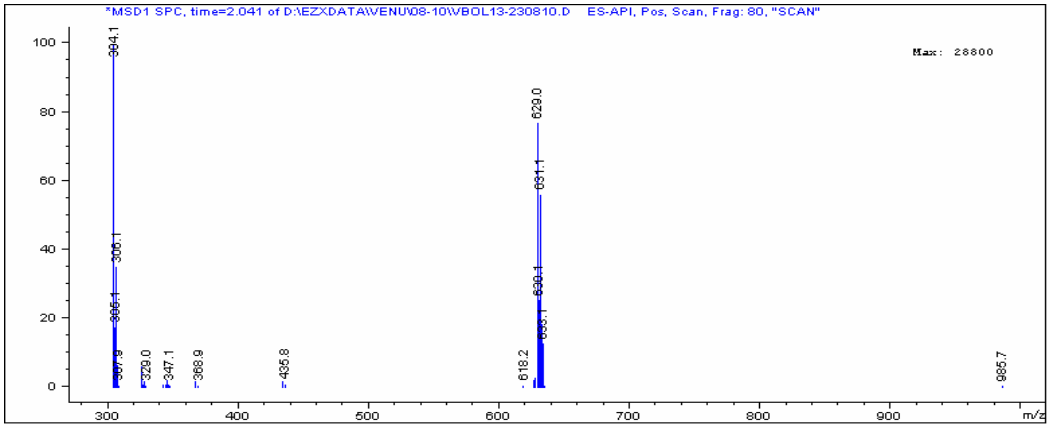
Compound 150



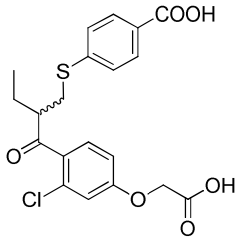


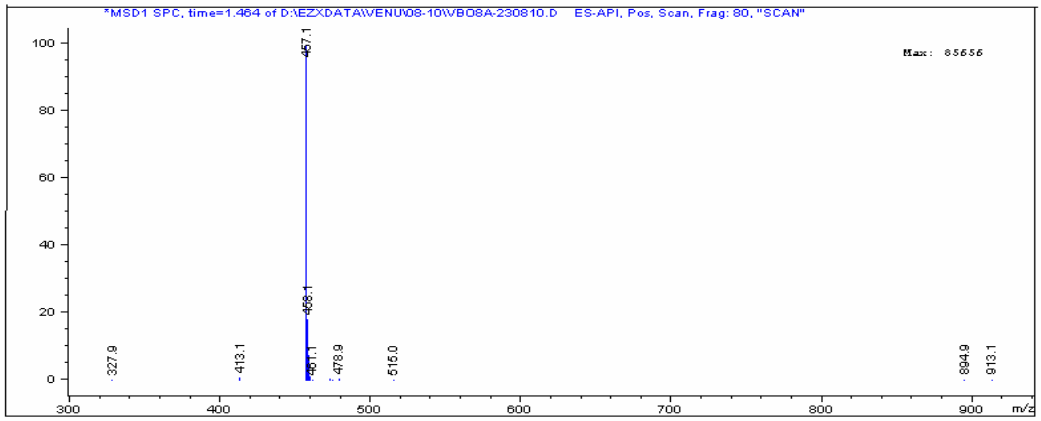
Compound 151





Compound 154





3.5 Notes and References

1 Corbett, P.; Leclaire, J.; Vial, L.; West, K. R; Wietor, J.-L.; Sanders, J.K.M.; Otto, S. *Chem. Rev.*, **2006**, *106*, 3652-3711.

2 Rowan, S. J.; Cantrill, S. J.; Cousins, G. R. L.; Sanders, J. K. M.; Stoddart, J. F. *Angew. Chem., Int. Ed.*, **2002**, *41*, 898-952.

3 Ludlow, R.F.; Otto, S.*Chem. Soc. Rev.*, **2008**, *37*, 101-108.

4 Corbett, P.; Sanders, J.K.M.; Otto, S. *Angew. Chem., Int. Ed.*, **2007**, *46*, 8858-8862.

5 Ludlow, R.F.; Otto, S. *J. Am .Chem. Soc.* **2008**, *130*, 12218-12221.

6 Sarma, R.J.; Nitschke, J.R. *Angew. Chem., Int. Ed.*, **2008**, *47*, 377-380.

7 Nitschke, J.R. *Nature* **2009**,*462*, 736-738.8 Mal, P.; Nitschke, J.R. *Chem. Commun.* **2010**, *46*, 2417-2419.

9 Campbell, V.E. ; de Hatten, X.; Delsuc, N.; Kauffmann, B.; Huc I.; Nitschke, J.R. *Nature Chem.* **2010**, 2, 684-687.

10 Wagner, N.; Ashkenasy, G. *Chem.–Eur. J.*, **2009**, *15*, 1765-1769.

11 Klekota,B.; Miller, B. L. *Tetrahedron*, **1999**, *55*, 11687-11697.

12 Goral, V.; Nelen, M. I.; Eliseev A. V.; Lehn, J.-M. *Proc. Natl. Acad. Sci. U. S. A.*, **2001**, *98*, 1347-1352.

13 Leclaire, J.; Vial, L.; Otto, S.; Sanders, J. K. M. *Chem.Commun.*, **2005**, 1959-1961.

14 ten Cate, A. T.; Dankers, P. Y. W.; Sijbesma R. P.; Meijer, E. W. *J. Org. Chem.*, **2005**, *70*, 5799-5803.

15 Sarma, R. J.; Otto S.; Nitschke, J. R. *Chem.–Eur. J.*, **2007**, *13*, 9542-9546.

16 Christinat, N.; Scopelliti R.; Severin, K. *Angew. Chem., Int. Ed.*, **2008**, *47*, 1848-1852.

17 Rodriguez- Docampo, Z. R.;. Otto,S. *Chem. Commun.*, **2008**, 5301-5303.

18 Orrillo, A. G.; Escalente A M.; Furlan, R. L. E. *Chem.Commun.*, **2008**, 5298-5300.

19 Escalente A M.; Orrillo, A. G.; Furlan, R. L. E. *J. Comb. Chem.*, **2010**, *12*, 410-413.

20 Sadownik, J.W.; Philp, D. *Angew. Chem., Int. Ed.*, **2008**, *47*, 9965-9970.

21 Inzgerman,. L.A.; Waters, M.L. *J. Org. Chem.*, **2009**, *74*, 111-117.

22 von Delius, M.; Geertsema E. M.; Leigh, D. A. *Nature Chem.*, **2010**, *2*, 96-101.

23 Otto, S. *Nature Chem.*, **2010**, *2*, 75-76.

24 Furlan R. L. E.; Ng, Y.-F.; Otto S.; Sanders, J. K. M. *J. Am. Chem. Soc.* **2001**, 1*23*, 8876–8877.

25 Roberts, S. L.; Furlan, R. L. E.; Cousins G. R. L.; Sanders, J. K. M. *Chem.Commun.* **2002**, 938-939.

26 Liu, J.; West, K. R.; Bondy, C. R.; Sanders J. K. M. *Org. Biomol. Chem.* **2007**, *5*, 778-786.

27 Bunyapaiboonsri, T.; Ramström, O.; Lohmann, S.; Lehn, J.-M.; Peng, L.; Goeldner, M. *ChemBioChem* **2001**, *2*, 438-444.

28 Cousins, G. R. L.; Poulsen, S.-A.; Sanders, J. K. M. *Chem. Commun.* **1999**, 1575-1576.

29 Bunyapaiboonsri, T.; Ramström, H.; Ramström, O.; Haiech, J.; Lehn, J.-M. *J.Med. Chem.* **2003**, *46*, 5803- 5811.

30 Lao, L.L.; Schmitt, J.-L.; Lehn, J.-M. *Chem.–Eur. J.*, **2010**, *16*, 4903-4907.

31 Shi B.; Greaney, M.F.; *Chem. Commun.*, **2005**, 886-888.

32 Shi, B.; Stevenson, R.; Campopiano D. J.; Greaney, M. F. *J. Am. Chem. Soc.*, **2006**, *128*, 8459-8467.

33 Nguyen R.; Huc, I.; *Chem. Commun.*, **2003**, 942-943.

34 Bhat, V. T.; Caniard, A. M.; Luksch, T.; Brenk, R.; Campopiano D. J.; Greaney, M. F. *Nature Chem.*, **2010**, *2*, 490-497.

35 Friedman, M.; Cavins, J. F.;. Wall, J. S. *J. Am .Chem. Soc.* **1965**, *87*, 3672-3682.

36 Addition of cysteine at acidic pH introduces the possibility of thiazolidinone formation, a recation studied for DCC by Wipf *et al* (reference 37). No thiazolidinone was observed, indicating that hydrazone formation is dominant in our DCLs at these stoichiometries.

37 Saiz, C.; Wipf, P.; Manta E.; Mahler, G. *Org. Lett.* **2009**, 15, 3170-3173.

Summary and Outlook

Dynamic combinatorial chemistry (DCC) is a unique approach to studying protein-ligand interactions which integrates the synthesis and screening of small molecule libraries into a single step. It achieves this by using reversible reactions to generate dynamic libraries that respond to template effects. The range of reversible reactions that can be used with biopolymer templates is limited, given that the thermodynamic criterion of reversibility has to be met under physiological conditions which preserve the structural and functional integrity of the template. A general research direction to establish DCC as a preferred tool in drug discovery is the development of new reversible chemistries which are compatible with biological targets.

In this thesis we have adopted a new strategy: using non-biological catalysts to extend the list of reversible reactions that can form adaptive DCLs under biocompatible conditions. Using Dawson's simple organocatalyst, aniline, we have generated adaptive acylhydrazone libraries that amplify isoform-specific inhibitors for two different GST enzymes. The molecular basis of this observed isoform-selectivity is not well understood and requires structural studies of the protein-ligand complexes. The protein-template induced amplification of a specific constituent of a DCL reflects not only the binding affinity of the selected member for the template, but also, in a very complex fashion, the affinities of all other members in the dynamic library. Our understanding of these thermodynamic minima and amplification effects at a systems level in biological DCLs is very preliminary at the current stage and has substantial scope for future research. To establish broad applicability of our approach in a drug discovery context, current projects in the Greaney group focus on using DCC to study more challenging protein-protein and protein- nucleic acid interactions. An immediate extension of our work currently underway uses aniline- catalysed bis acylhydrazone DCLs to identify bivalent inhibitors of GST.

In Chapter 3 of this thesis, we have established that thiol conjugate addition and hydrazone formation exhibit very high levels of orthogonality when combined in a single DCC experiment. The DCLs are formed in aqueous solution and are easily

controlled by simple pH change, undergoing cyclical de- and re-activation according to pH regime.

The field of DCC provides an entry into the nascent field of systems chemistry, defined as a study of complex molecular networks. Orthogonal reversible chemistries can be used to design oscillatory molecular networks which may serve as synthetic mimics of metabolic and gene regulatory circuits. Simultaneous multi-level dynamic libraries using multiple reversible reactions in a single system offer tremendous scope for generating newer network topologies. Enmeshing the principles of network science with complex molecular mixtures would be a giant step towards realising ‘minimal synthetic cells’ which will play a role in understanding the mysteries of cell biology.